

Genetic revolution overdue at the USDA

Agricultural research is being neglected by the US federal government, despite its clear relevance to economic prosperity, social development and environmental stewardship.

The United States Department of Agriculture (USDA) has, more than most agencies of the US government, a history of achievement that demands respect. The department's outreach to farmers, particularly in the aftermath of the dustbowl and depression of the 1930s, brought cheap food to America's kitchen tables and helped build the mid-western plains into the bread-basket of the world.

But, 45 years after Watson and Crick transformed our understanding of the nature of living things, the USDA's approach to scientific research remains firmly rooted in that previous golden era. The structure of its \$1.5-billion annual research programme carries little recognition of the genetic revolution, even as the first agricultural products of that revolution sweep across the rice, soybean, wheat and corn fields of the nation. The importance of the change is better recognized on Wall Street than at either end of Pennsylvania Avenue. But, as the potent symbiosis between the National Institutes of Health (NIH) and the US biotechnology sector amply demonstrates, government can vastly accelerate the progress of biotechnology by supporting basic research.

In plant science, this isn't happening: the USDA's support for basic research is flat, its grants are each too small, and it allows overhead rates so miserly as to effectively discourage university participation in its programmes (see page 210). The USDA should be leading a great investigation of plant genetics and a drive to sequence the genomes of both scientifically interesting and economically significant plants. But it has instead fallen to the National Science Foundation — at the rude insistence of one senator, Christopher Bond (Republican, Missouri) — to initiate the new plant genome initiative (see *Nature* 388, 309; 1997).

Part of the problem is political. The USDA has a sprawling empire of special interests to tend. Agricultural research has never been fenced off from those interests, as the NIH is from the rest of the

Department of Health and Human Services. To judge from the track record of the respective arrangements, it ought to be.

USDA also has a strong tradition of operating its own research centres and directly supporting local agricultural colleges. Sadly, but perhaps inevitably, these institutions have never taken kindly to the allocation of research money on a fully competitive basis, fearing that too many of the grants would end up at major research universities. For decades, farmers, agricultural colleges and congressmen from agricultural districts (who naturally dominate the committees controlling USDA) have often found themselves fearing scientific excellence and the advent of molecular biology, when it would have been in their best interests to embrace both.

Those fears are waning now and most of the agricultural schools have warmed to genetics, genomics and competitive peer review. Agribusiness has sought to persuade Congress of the value of basic research to the entire agricultural enterprise. It has found allies, such as Bond and Senator Richard Lugar (Republican, Indiana), chair of the Senate Agriculture Committee, who has proposed and enacted a law that would inject mandatory funds — money not subject to yearly haggling by the appropriations committees — into agricultural research.

This manoeuvre is proving difficult to execute, however, and it isn't clear that the White House will go out of its way to secure its success. This administration has, in fact, shown little interest in agricultural research. Yet there exists a serious imbalance between its strategic significance and the resources that are made available for it. It is true that the public can't see the need for agricultural research so vividly as it sees the need for health research, but the requirement is real enough, if the world is to feed itself without destroying its natural environment. When Neal Lane is confirmed as President Bill Clinton's new science adviser, as he will be very shortly, he should treat the matter as an urgent priority. □

A Wellcome break for science

The significant boost in funds for UK science is a triumph, above all, for the Wellcome Trust.

Coming hard on the heels of the expressed determination of President Bill Clinton to provide more support for science, this week's announcement by the British government of a major boost in science funds (see page 209) sends a positive signal to researchers and governments everywhere. Moreover, it has the additional strength of being a firm three-year commitment, rather than subject to the annual wrangling that occurs in the US Congress.

The message is that long-term commitment to fundamental science is essential if you want to help the quality of your citizens' lives to improve. That message appears to have been successfully driven home to the people who really mattered — UK Treasury ministers — by learned societies, lobbyists, such as Save British Science, pharmaceutical companies and, by no means least, by the Office of Science

and Technology on behalf of the research councils. As we go to press, the signs are that the university funding councils will at the least continue to be well supported.

But the Wellcome Trust's role has been crucial. With the trust's annual funding of science in the region of almost £300 million (US\$490 million), it was all too tempting for the government to try to offload some of the responsibility for increased support for basic research. To her credit, the trust's recent head, Dame Bridget Ogilvie, would have none of it. Partly as a result, the trust has been able to play a central role in leveraging increased government support for world-class science in Britain (it has no obligation to restrict its activities there). Research itself will be immeasurably healthier as a result. □



UK provides 'step increase' in funding for its science base

[LONDON] Britain's Labour government made good on pre-election signals this week when it announced that, egged on by the Wellcome Trust, it has agreed to inject an extra £700 million (US\$1.14 billion) into the country's science base over the next three years.

The decision, which will mean an increase of 7.3 per cent in research funding through the six UK research councils in the financial year that starts next April, has emerged from the comprehensive spending review launched by the government when it came to power last summer.

Extra government funding is being complemented by a further £400 million from the Wellcome Trust over the same period, making a total of £1.1 billion of new money. Of this, £600 million — shared jointly by the government and the trust — will go to the construction and refurbishment of university laboratories, a priority identified last year by the Dearing report on higher education (see *Nature* **388**, 413; 1997).

In addition, the government is to make an extra £400 million available to the research councils to meet the current and capital costs of new project funding in priority areas "like life sciences". And Wellcome is to provide £100 million towards the cost of a new synchrotron source (see right).

According to the Department of Trade and Industry, which is responsible for the Office of Science and Technology (OST) and for the budget of the research councils, the result will be a 23.8 per cent cash increase over its 'baseline' provisions for science in 1998–99. That will mean a total of £1.67 billion by the financial year 2001–02 (in real terms, the increase will be 14.8 per cent).

"This is a step change for British science and engineering that will transform the funding of the science base after many years of neglect, and set us on a strong course for investment in the future," says Margaret Beckett, President of the Board of Trade, and the cabinet minister responsible for science.

Much of the new money will be to develop what John Cadogan, the director-general of the research councils — and a key figure in putting the science community's request for higher funding to the government — calls 'post-genome science', a broad category that covers all aspects of the exploitation of data emerging from human and other genome sequencing efforts.

Cadogan has argued strongly that, as this is a field in which Britain already has a strong competitive advantage, the future health of its economy depends on its success in finding

profitable applications for gene sequencing data, in particular for the prevention, diagnosis and treatment of disease. The same argument has been put by much of the UK pharmaceutical and biotechnology industry.

Not surprisingly, the government's endorsement of this strategy has been warmly welcomed in the biomedical research community. "It is very exciting for us," says George Radda, secretary of the Medical Research Council.

Both Cadogan and Mike Dexter, the new director of the Wellcome Trust, emphasize that the task of understanding and exploiting genomic data will involve a wide range of research disciplines, from computer technology to fundamental physics. "This is a unique opportunity [for Britain] on the threshold of the next millennium," says Dexter.

There was a universal welcome on Monday for the new funding. "This has been a good day for British science," said John Mulvey, secretary of the pressure group Save British Science, which has long pushed for increases of this order of magnitude. "We are saying 'well done' to the OST, while the actions of Wellcome have been magnificent in ensuring the repair of university equipment."

Any overall assessment of the impact of the spending review will depend on the allocations to universities, which had not been announced as *Nature* went to press. But the indications were that, although the universities might be required to seek higher overhead repayments from research councils to cover their costs, threats to the 'dual support' system have not materialized (see *Nature* **393**, 607; 1998).

David Dickson

Wellcome secures new synchrotron source

[LONDON] Christmas has come early this year to the 3,000 British users of synchrotron radiation with the announcement that the Wellcome Trust is to provide £100 million (US\$163 million) towards the cost of a new synchrotron facility.

The figure represents well over half of the anticipated construction costs of Diamond, the 3-GeV X-ray synchrotron machine planned as a successor to the world's first dedicated X-ray source, the 2-GeV Synchrotron Radiation Source at the Daresbury Laboratory in Cheshire.

Britain's research councils are expected to agree shortly to put up the remaining sum. They had been baulking at the prospect of having to finance the total construction cost and had been frantically seeking other possible sources of funding — including even resorting to bank loans.

The Wellcome announcement came as part of a broad £1.1 billion package of additional support for science announced on Monday (13



Dexter: an 'essential' tool.

July) by Margaret Beckett, the cabinet minister responsible for science (see above).

The trust's interest in contributing towards the cost of the facility had been known since last November, when it announced that it would give £10 million towards the project (see *Nature* **389**, 318; 1997). The move was seen as a bid to catalyse decision-making in government circles.

It appears to have succeeded. Monday's unexpected announcement that Wellcome had decided to increase its contribution by an order of magnitude has been met with surprise and delight in the synchrotron research community, whose members range from

structural biologists to materials scientists.

"I am enormously pleased," says Guy Dodson, professor of chemistry at the University of York. "It is a remarkable commitment which means that the next generation of biologists and physicists can now be certain of having a level playing field in comparison with facilities available in the United States and elsewhere."

Mike Dexter, the new director of the Wellcome Trust, points out that the ability of synchrotron radiation to resolve the structure of small molecules has made it an essential tool for structural biologists.

This is a field in which Britain has many top-level researchers, including John Walker, last year's Nobel prize winner for chemistry (see *Nature* **389**, 771; 1997). "We are determined to see that we keep these scientists in the United Kingdom, and ensure that they have the resources necessary to continue competing with the best in the world," says Dexter.

D.D.

Fusion machines 'breach treaty and open weapons risk'

[WASHINGTON] Both the National Ignition Facility (NIF), now under construction at the Lawrence Livermore National Laboratory in California, and the Laser Megajoule project planned near Bordeaux, France, will violate the Comprehensive Test Ban Treaty (CTBT) and could open the way for the development of new, 'pure fusion' nuclear weapons, according to a study published this week.

The Institute for Energy and Environmental Research (IEER), a technically respected group based in Takoma Park, Maryland, calls on parties to the CTBT to issue a ruling on the treaty status of inertial confinement fusion research, which would be conducted at the two facilities. The IEER, which is consistently critical of US nuclear weapons policy, wants both projects halted.

The IEER's report, *Dangerous Thermonuclear Quest*, says that an operational NIF will lead to strong political demand for so-called 'pure fusion' weapons (see *Nature* 387, 439; 1997). These would yield fusion explosions without any need for the fission explosions that trigger fusion in existing thermonuclear weapons.

Both NIF and the Laser Megajoule will generate tiny thermonuclear explosions of a few kilogrammes TNT equivalent, triggered by massive and immobile banks of laser beams. "NIF will not, by itself, lead to pure fusion weapons," says Arjun Makhijani, a plasma physicist who heads the IEER. "But it could play a crucial role by enabling the design of targets for other driven systems, which could be miniaturized."

Makhijani also argues that NIF will create political pressure for the development of pure fusion weapons. "We must prevent the scientific feasibility of these weapons being established, because once it is, the pressure to build them will become irresistible," he says.

David Crandall, director of the NIF office at the Department of Energy, says his department's non-proliferation office and the 'Jasons', a group of scientists who advise the government on complex issues, have confirmed that the machine poses no proliferation risk. Crandall says the signatories to the CTBT accept the definition of a nuclear explosion contained in the earlier Non-Proliferation Treaty (NPT). The NPT specifically permits inertial confinement fusion research.

Makhijani says the CTBT bans all nuclear explosions. "If they wanted a specific exemption for this, they should have asked for it."

The NIF has been under construction for twelve months, has strong Congressional support, and, as Crandall confirms, there is no chance of it being halted in response to the report.

Colin Macilwain

Plant scientists want focus on the best and brightest

[WASHINGTON] Scott Poethig, a geneticist at the University of Pennsylvania, Philadelphia, who studies the genes controlling the switch from juvenile to adult development in plants, doesn't need to do an in-depth analysis of research statistics to figure out that his branch of science is getting a raw deal from the federal government.

As a grantee of the United States Department of Agriculture (USDA), Poethig's current grant is worth \$66,000 and lasts for just twelve months. Poethig's wife, Maja Bucan, a behavioural geneticist at the same university, works with mice instead of plants. She is funded by the National Institutes of Health (NIH), where the typical grant, according to agency officials, provides \$240,000 a year and lasts for four years.

Last week, Poethig took his case to the august halls of the USDA in Washington. At a meeting on 9 July, called by the department to solicit advice on how it should manage a planned new research initiative, he told top science officials at the department to find the best basic research, and support it to the hilt.

"My suggestion is simple," he said. "Make the USDA more like NIH. Fund the best research, not just research on economically useful plants." The brightest scientists are working to understand the basic nature of biological systems, Poethig says; the USDA would best serve US agriculture by supporting them.

Poethig reflects the views of many plant scientists at US universities. Interest in their field is poised to explode: in two weeks' time, for example, private benefactors will announce that they have raised \$150 million to build a new plant-science institute in St Louis, Missouri (see box). But according to US government figures, support for research and development at the USDA, the main federal sponsor of plant science, slipped from

\$1.3 billion in 1992 to \$1.2 billion last year. USDA support for basic research also fell.

The problem runs deeper than simple resources. Despite a wide consensus in the United States that peer-reviewed, extramural grants are the most efficient way to support good science, less than 10 per cent of USDA research money goes on such grants. For at least two decades, various agricultural constituencies, allied with key members of Congress, have fought to resist change that they believe would shift resources from local research stations and agriculture schools to research universities. "I find this very painful," says Lou Sherman, head of the biology department at Purdue University, Indiana. "It has forced one area of science to be less good than it could have been."

When the USDA started funding competitive grants under the so-called National Research Initiative (NRI), Congress specified that these grants could allow universities to receive overhead payments of only 14 per cent of the value of the grant, compared with payments of as much as 50 per cent on grants from other government agencies.

Sherman calls this a "poison pill", which dissuades universities from chasing the money. University administrators, he says, conclude that agricultural research is not a good area to pursue; they then hire fewer plant and soil scientists. "Congress has hurt agricultural research by making it feel like a poor relation," says Sherman.

Even the chief scientist at the USDA, Michael Roberts, who also holds a position at the University of Missouri, concurs with these complaints. "If we're not careful, the NRI is going to be marginalized," he says, because its grants are too small to sustain researchers for long. Typical NRI grants are worth \$100,000 and last for two years. The initiative was originally authorized by Con-



Stunted growth? University agricultural researchers, such as these at the University of Florida studying the effects of rising CO₂ on crops, can feel badly treated compared to their colleagues.

AP/UNIVERSITY OF FLORIDA

gress in 1990 as a \$500-million-a-year programme, but annual funding for it has never exceeded \$100 million.

Advocates of agriculture research in Congress, such as Senator Richard Lugar (Republican, Indiana), chairman of the



Lugar: seeks to raise research spending.

Senate Agriculture Committee, are well aware of this shortfall. The new programme discussed at last week's meeting, the Initiative for Future Agriculture and Food Systems, results from a bill proposed by Lugar and signed into law by President Bill Clinton on 23 June. The bill would use mandatory funds, previously used in the food stamp programme, to support \$600 million of new agricultural research over five years.

But the proposal is already in trouble on Capitol Hill. The House Appropriations Committee, in a bill marked up only days after Clinton signed the legislation, expressly blocks the \$120 million allocated for the measure in the next financial year. The Senate appropriators allowed the spending, leaving the fate of the initiative to be determined in September, when the two chambers reconcile their budget proposals.

Even if the money is forthcoming, some fear that it will be spread too thinly across the activities — research, extension and education — it is supposed to support. Although the legislation specifies some priorities for grants, including genomics, biotechnology and food safety, a dizzying array of research interests feel entitled to a share of the pot.

As speakers at the 9 July meeting made clear, everyone wants a piece of the action, from international economists and organic farmers to environmentalists and soil scientists. The last group's advocate, Karl Glasener, at least introduced some humour: "In short, soil scientists would like everyone to stop treating soil like dirt," he says.

"The thing I worry about most is that it'll be divided up between every interest group, so that nothing that great will come out of it," says Kelly Eversole, a lobbyist for the American Cornrowers' Association, which, along with other growers' groups, wants money to go to large collaborations in areas, such as genomics, that will raise agricultural yields. "We want large, multi-institutional, well-organized projects," says Lyle Roberts of the American Soybean Association.

Eileen Kennedy, deputy under-secretary for research at the USDA and the agency's senior science official, says the department has introduced more competition into its

Monsanto backs \$150m plant science centre

[WASHINGTON] The US life sciences company Monsanto is linking up with a charitable trust to create an independent \$150 million plant science institute in St Louis, Missouri, that is intended to become an international centre of excellence for interdisciplinary plant research.

Plans for the new centre, strategically placed at the heart of America's agricultural mid-west, are due to be announced on 31 July by former president Jimmy Carter. It will operate as a joint venture between the Missouri Botanical Garden, Washington University in St Louis, the University of Missouri at Columbia, and Monsanto.

Although Monsanto will contribute cash, land and tax credits worth over \$80 million, it says that it will not lay any claim to intellectual property generated at the institute, which is expected to attract research support from government, industry and private foundations.

A search committee to find a director for the institute by the end of the summer is being chaired by Peter Raven, director of the Missouri Botanical Garden. "We're talking to some of the best plant scientists in the world," says Sam Fiorello, an assistant to the president of Monsanto.

"This is an opportunity to develop for our region a

leading centre that will bring together a critical mass of outstanding research," says Mark Wrighton, chancellor of Washington University. "We're depending a lot on the recruitment of an outstanding director."

The partners plan to spend \$45 million on constructing a 200,000 square foot building to house the institute, opposite Monsanto's St Louis headquarters, and a further \$15 million to equip it.

The Danforth Foundation, a St Louis-based charitable trust which will probably give its name to the new institute, has promised to contribute \$6 million a year for ten years to operate the centre. It is expected to employ 15 principal investigators and 105 staff in total.

"The scope and orientation of the institute will be something like a Max Planck or Pasteur Institute," says Raven. "What is unique is the combination of private funding with excellent research organizations already based here in St Louis."

He adds that it is "quite remarkable" that Monsanto is putting so much money into the centre "when they expect to exercise no control", adding that the corporation "would have had no difficulty putting the money into its own research". Monsanto is in the process of merging with American Home Products, a consumer-goods

corporation, but has said that the life sciences operation of the combined group will be based at St Louis.

Each of the four partners will have one representative on the institute's governing board, which will be chaired by William Danforth, chair of the board of trustees at Washington University. William is the brother of John Danforth, the former Missouri senator and chair of the Danforth Foundation. Roy Vagelos, the former chief executive of Merck, is the first of two outside directors who will join the board.

Raven, William Danforth and Virginia Weldon, a recently retired Monsanto executive, dreamed up the idea for the centre on their way to a National Research Council meeting at Irvine, California, in February last year. "I had a vision that this region should be strong in plant biology," says Danforth, adding that the new centre will be "embedded in a community that has a lot going on" in the discipline.

Raven says the great gains in agricultural productivity of the past 50 years have often been made at the expense of the planet's productive capacity, and that the new institute will help to establish ways of raising productivity while preserving topsoil and biodiversity. "We've got to learn to live off the interest, not the principle," he says.

C. M.

intramural research programmes, and is considering larger grants, of \$250,000 or more, for university researchers. Kennedy says that politicians' reluctance to support agricultural research reflects the public's belief that, with food so cheap in the stores, the sector's problems have been solved.

She thinks that "groups that have traditionally been locking horns" — farmers, agribusiness interests and universities — are united behind the initiative. "I don't sense any opposition to agricultural research" in Congress, she says. "It is just that they have a smaller pot of money than they'd like."

No one can predict how much of the \$120

million will be delivered when the dust settles on the USDA budget in the autumn. If any money does appear, Kennedy will issue an immediate request for proposals, and awards will be made early in the new year.

Senator Lugar's initiative will not enable a fully fledged revival of agricultural science in the United States, and agriculture schools will not soon match the opulence of many academic health centres. But, together with the National Science Foundation's plant-genome initiative (see *Nature* 390, 539; 1997) and growing private investment, it may herald a modest revival in an undervalued branch of US science.

Colin Macilwain

UK court gives green light to trial of modified maize

[LONDON] The UK campaign for a moratorium on commercial genetically modified crops suffered a setback last week when a farmer failed in his bid to halt a trial of genetically modified maize planted next to his crop of organically grown sweetcorn.

The farmer, Guy Watson, has appealed against the decision by the High Court in London to allow the experimental trial in Devon to continue. He is being backed by the environmentalist group Friends of the Earth and the Soil Association, which certifies his organic produce (see *Nature* 394, 8; 1998).

Watson says he hopes the trial, by the company Sharpes Seeds, can be stopped before the flowering season starts next week. He argues that pollen from the genetically modified maize could pollinate his own crop, which would lose its organic certificate.

But the court ruled that cross-pollination was unlikely, as Watson's sweetcorn is two kilometres away from the genetically modified maize. The decision was based on advice from the government's advisory committee on releases to the environment (ACRE), whose members concluded that cross-pollination was "likely to be zero".

Despite the decision, and the scepticism of most ACRE members about any significant threat to human health from genetically modified crops, the British government remains under pressure to adopt a moratorium on their commercial introduction.

Last week, for example, English Nature, the government's official conservation advisory body, reiterated a call for a three-year moratorium on their commercial release "until current research on their potential effects has been completed and analysed."

"We are certainly not against the development of genetically modified organisms," says Brian Johnson, English Nature's adviser on issues affecting genetic modification. "But we are concerned that their introduction should take place carefully and against a

background of sound knowledge."

Chris Haskins, chairman of Northern Foods and Express Dairies, said a three-to-five-year moratorium was needed to give the public time to regain confidence in government scientific opinion.

The question of public confidence in the government's scientific advisers attracted widespread media attention last week when Friends of the Earth pointed out that six of ACRE's thirteen members had interests in companies or organizations that had applied to the committee for licences for trials.

The environmentalist group also pointed out that no application for a trial licence has yet been recommended for refusal. "How can people have confidence in the government advisory panel when so many members have close financial links to the biotech industry?" asks Adrian Bebb, a food campaigner with Friends of the Earth.

But John McLeod, an ACRE member and director of the National Institute for Agricultural Botany, which is conducting the Devon trial on behalf of Sharpes Seeds, says that relevant ACRE members leave the room and do not look at paperwork when the committee discusses licensing applications from organizations with links to ACRE members.

McLeod says this conflict of interest is not of his — or his colleagues' — making. They are appointed to ACRE on the basis of their expertise in assessing the risks of genetically modified crops, but must apply to the same committee for licenses for their own research.

The government has not yet responded officially to the controversy, except to say that it takes the issue seriously and will "shortly" issue a response to public concern about the ethical implications of commercial genetically modified crops. Some changes are already underway, for example 11 ACRE members will be replaced next year, with more members being drawn from the public.

The move is welcomed by Julie Hill of the Green Alliance, the only ACRE member from an environmentalist group, who argues that many members, although of the highest professional integrity, "are 'pro-technology', which colours their view on how they judge the acceptability of a risk."

Many members say the committee needs more experts with experience of research into the implications of genetically modified organisms. Andrew Watkinson, professor of ecology at the University of East Anglia, and an ACRE member until 1993, says "The committee handles some very technical issues in molecular biology, which environmentalists and ecologists would find difficult to comment on."

Ehsan Masood

Britain ignores warning on privatization of defence research

[LONDON] The British government appears set to press ahead with plans to introduce some form of private management of its Defence Research Agency (DERA), against the advice of the House of Commons defence select committee.

In a long-awaited review of Britain's defence policy — the Strategic Defence Review, published last week — the government does not rule out further private-sector involvement in defence research.

The review was released a day after the defence select committee, in its own report on DERA, warned the government against inviting private companies to manage DERA and its affiliated agencies, such as the chemical and biological defence establishment at Porton Down.

But the government's review fails to provide any additional detail on DERA's future beyond what is known already. Ministers have said the agency's fate lies somewhere between total privatization and the status quo, in which it remains part of the Ministry of Defence and is run by the government.

One set of alternatives would involve allowing private companies to manage DERA, or its affiliated organizations, and activities. This arrangement would be similar to other government-owned but privately managed laboratories, such as the National Physical Laboratory in Teddington.

A second alternative is a 'public-private partnership', in which a facility or service is paid for by the private sector, with the government paying for the cost of use. The recently announced supercomputer service for UK academics will be provided in this way (see *Nature* 394, 114; 1998).

A privately run DERA, the select committee report says, could damage DERA's relations with industry. At present, the agency is seen to be impartial in the work it performs and the advice it gives, both to industry and government. Private-sector management could change that, as the agency will continue to compete with industry for the Ministry of Defence's research programmes.

International research collaboration could also be a casualty of increasing private-sector involvement in DERA, according to the report. Committee members say the US government, for example, is keen on transatlantic research cooperation.

The members also conclude in the report that international collaboration "would be seriously at risk if privatization of DERA proceeded." Julian Brazier (Conservative, Canterbury), a member of the select committee, says that "the committee's message is very clear: leave DERA alone."

Ehsan Masood



Gore's satellite may fill role of lost SOHO

[WASHINGTON] An Earth-viewing satellite that few Earth scientists want may be of more use to space physicists stunned by the apparent loss of the US–European Solar and Heliospheric Observatory (SOHO) (see *Nature* 394, 5; 1998).

The US space agency NASA last week asked for proposals for the Triana spacecraft, suggested by US Vice-President Al Gore in March as a way of inspiring the public with live pictures of Earth taken from space (see *Nature* 392, 220; 1998).

The \$50 million satellite, to be launched in late 2000 (when Gore is expected to be campaigning for the presidency), would be parked in a stable orbit 1.5 million kilometres from Earth in the direction of the Sun.

According to a reference mission design provided by NASA, the only instrument on board would be a colour camera capable of returning a new image of the whole Earth every three minutes. The best resolution would be 14 kilometres; this is poor by the standards of the remote-sensing satellites now in Earth orbit, which can see details as small as a few metres across.

Earth scientists have been largely indifferent to Gore's idea, or have grumbled, in private, that it may take funds from more legitimate peer-reviewed projects. Most of the scientists who answered an informal NASA request for ideas for Triana in April proposed adding something else to the satellite, such as solar radiation sensors or a more capable camera.



Earth in view: Gore's announcement (above) of his Triana project has met criticism in Congress.

Such extras may in fact be necessary to save the project. Appropriators in the US House of Representatives have blocked funding for Triana until NASA can show — based on the proposals it receives by the end of August — that the idea has support in the scientific community or private sector.

"NASA must demonstrate for the [appropriations] committee that the agency has a plan for a public-private, peer-reviewed mission, which has resulted from a competitive process," wrote the lawmakers in a report accompanying NASA's 1999 spending bill. The chairman of the House subcommittee on space, Dana Rohrabacher (Republican, California), has dismissed Triana as a "\$50 million politically correct screensaver".

NASA's solicitation encourages proposers to embellish the basic single-camera concept, as long as they supply the additional funds. The fact that Triana is destined for the same unusual orbit as SOHO has drawn the attention of NASA managers working on contingency plans should the \$1 billion solar observatory be declared a loss.

George Withbroe, who heads NASA's solar and space physics programme, says it may be several months before SOHO's fate is known, as its solar panels may slowly turn to catch enough sunlight to repower its batteries. Although engineers are uncertain this will work, or what condition SOHO would be in after a long dormant period, Withbroe says "I'm not going to write it off yet."

But, to preserve the option of launching some solar-observing instruments in time to monitor the solar sunspot maximum expected to begin next year, officials from NASA and the European Space Agency have already started looking at quick, cheap recovery plans. Triana's fast track to launch makes it especially appealing.

The proposed spacecraft is much smaller than SOHO, but Withbroe says it may be able to piggy-back some high-priority instruments, such as a coronagraph, if necessary.

NASA is looking at other options, such as speeding up the STEREO solar-imaging mission now planned for launch in 2004.

US and European space managers will meet this month for further discussions on recovery plans for SOHO. **Tony Reichhardt**

Biologists recommend scrapping NASA's research on crystals

[WASHINGTON] A panel of biologists reviewing the US space agency NASA's life science research programme has called for an end to protein crystallography experiments in space — one of the highest-profile research activities planned for the International Space Station.

The committee of the American Society for Cell Biology (ASCB) also says that research into "basic animal and plant cell and developmental biology cannot be used to justify a space mission".

The panel of seven scientists was chaired by Donald Brown, a developmental biologist at the Carnegie Institution of Washington, DC. It was asked by ASCB president Elizabeth Blackburn of the University of California at San Francisco to develop a view on life science research in space in the light of opposition voiced by some biologists.

The resulting report was passed unanimously by the society's governing council last month, and includes some of the most pointed criticism ever levelled at NASA research by a scientific body.

The panel called space-based experiments to investigate gravity-sensing organs "premature" until more basic research is done on the ground. The group saw potential value in NASA ground-based plant research, but said that space experiments in general should have to justify their great cost, considering that the space station "will be the most expensive and inflexible laboratory ever built".

The harshest words were reserved for protein crystal growth in space, which NASA-funded researchers claim is useful in drug design. "No serious contributions to knowledge of protein structure or to drug discovery or design have yet been made in space," wrote the panel. "This committee recommends that no further funds be spent on crystallization of proteins in space."

Panel member Ursula Goodenough of Washington University in St Louis, a past president of the ASCB and a space station critic, said the group strayed somewhat from its territory of cell biology in judging the NASA crystallography work. But she

says that the presence on the panel of experts such as protein researcher Stephen Harrison of Harvard gave the committee confidence in its conclusion. Other members were Anthony Mahowald of the University of Chicago, Elliot Meyerowitz of the California Institute of Technology, Christopher Somerville of Stanford University and Carnegie, and Andrew Staehelin of the University of Colorado.

Tim Roemer (Democrat, Indiana), the space station's chief opponent in the US House of Representatives, was due to hold a press conference this week to publicize the report as part of his annual attempt to cancel the station when it comes to a House vote. That effort is almost certain to fail, just as a similar attempt failed in the Senate last week by a vote of 66 to 33.

Although the ASCB report is unlikely to have any real effect on the fate of the station, which begins construction this year, it is embarrassing for NASA, whose administrator, Daniel Goldin, has enlisted biologists' support for space research. **T.R.**

Israel split on rights to genetic privacy

[JERUSALEM] Leading officials in the Israeli government are reacting cautiously to a bill proposing stringent restrictions on the use of genetic information. The bill was passed last week by the Israeli parliament's science committee.

It was submitted by Knesset member Meir Shitrit, based on proposals drawn up by George Annas, professor of public health at Boston University School of Public Health, who has been closely involved in similar legislation in the United States (see *Nature* 384, 202; 1996).

The bill would grant individuals considerable control over genetic material taken from them for either medical tests or research purposes, and would forbid employers and insurance companies from discriminating on the basis of a person's genetic profile.

But Bracha Rager, chief scientist at the Ministry of Health, calls Shitrit's bill "problematic", arguing that it could prevent doctors from performing important medical tests, and might also act as an obstacle to medical research.

Shitrit complains that, at present, Israel lacks any controls over the way genetic information is handled. "There is no supervision, no licensing; no one explains to people about the implications of genetic tests, and there is no genetic counselling," he says.

His original bill contained a sweeping provision making all genetic material the property of the donor, who would have the right to demand that such material be destroyed at any time he or she chose.

This provision has been removed from the version that will go to the full Knesset for its first reading. But other provisions would still grant citizens considerable control over any of their genetic material that is likely to be obtained by doctors and researchers.

Shitrit argues that such controls are necessary to prevent the genetic information from being misused. But Rager contends that it is could be detrimental to patients.

For example, she says, someone who has provided genetic material to be tested for either of two genetic markers for a form of cancer, and who then asked for the material to be destroyed, could miss out on a vital test if a third marker were subsequently discovered.

Rager agrees with the provision in Shitrit's bill that genetic material donated for research purposes should be anonymous. But she adds that researchers should still be permitted to know certain basic features of the donor, and fears that the proposed law will prevent scientists from obtaining even these essential data.

Gali Ben-Or, an assistant deputy attorney-general who has been working on the issue of genetic property for the Ministry of

Justice, also believes that Shitrit's proposal is too broad. "Our position is that the bill is not constructed to take into account current legislation," Ben-Or says.

She argues that Israel already has laws dealing with similar issues, such as patients' rights and medical privacy. Ben-Or says it would be better to amend these laws to cover the new questions raised by genetic testing, rather than pass an entirely new law.

She points out, for example, that Israel's national health law guarantees health insurance to all citizens, and therefore there is no fear that an Israeli who has inherited a risky gene would not be able to get health insur-

ance. She also said that there might be some forms of employment, such as high-risk jobs involving, for example, exposure to radiation, where a person's genetic predispositions might be a legitimate consideration.

But not everyone agrees. Asa Kasher, who holds a chair in the professional ethics and philosophy of practice at the University of Tel Aviv, and is chairman of the National Research and Development Council's committee of ethics of science and intellectual property, also fears that some of the provisions of Shitrit's bill are too broad. But he accepts that special legislation is necessary.

Haim Watzman

Swiss databank to start charging for use

[MUNICH] The Swiss-Prot protein sequence databank, one of the world's most widely used reference sources on proteins, is introducing a system of licensing for private companies to ensure stable financing and help fund an expansion of its activities. Academic researchers will still be able to use the databank without charge.

Swiss-Prot is a joint enterprise between the newly created Swiss Institute of Bioinformatics and the European Bioinformatics Institute, the Cambridge-based outstation of the European Molecular Biology Laboratory (EMBL).

It is particularly valued by scientists because of the information added to simple translations of publicly available nucleotide sequences through the detailed annotation applied by the database's staff.

Until now, Swiss-Prot has been fully supported by public funding, and has provided a free service to both industry and academic researchers. But it hit a major funding crisis in 1996 when two major grants from the Swiss government were not renewed (see *Nature* 381, 266; 1996). It only managed to avert closure through a campaign for interim support, and is currently running on support from Swiss funding agencies.

Amos Bairoch, who founded Swiss-Prot in 1986, says it resisted buy-out offers from several commercial companies because he and his colleagues were concerned that commercial interests might be reluctant to continue maintaining open access to academic researchers.

Instead they are planning to develop a more stable funding strategy under which Swiss-Prot would receive half of its funding from public sources and half through licence fees from private companies.

Bairoch helped to set up the Swiss Institute of Bioinformatics from five Geneva-based bioinformatics research

groups, including those already involved in Swiss-Prot. It is independent of the university system.

As such, it is entitled to receive direct funding from the government without depending on university grant money. The government has already indicated that it considers Swiss-Prot a funding priority.

Bairoch says he also encouraged the creation last November of an independent Geneva-based company, GeneBio, which, in addition to its core business of developing high-quality specialized protein databases, will handle the administration of Swiss-Prot's licensing scheme.

Fees will range from US\$2,500 for small start-up companies to US\$90,000 for large companies. Pharmaceutical companies are used to paying this level of fee for databases, says Bairoch.

Fotis Kafatos, director-general of EMBL, emphasizes that the decision to impose licence fees will have no repercussions for nucleotide databases at the European Bioinformatics Institute, which are repositories of raw data. He points out that Swiss-Prot licensees will be paying for the "very labour-intensive expert curation".

The new funding system will allow Swiss-Prot to expand. At present the database includes 73,000 proteins, but an additional stockpile of 150,000 'virtual proteins' — those whose sequence is deduced from nucleotide information, but have not been functionally analysed — are waiting in the wings to be worked on.

Swiss-Prot staff stress that academics will continue to be exempt from charges. 'Grey' academic research areas involving small university-based start-up companies or industrial collaborators will be negotiated on a case-by-case basis, says Bairoch, adding that Swiss-Prot does not wish to interfere with academic enterprise.

Alison Abbott

Varmus agrees on need for more public input on NIH research...

[WASHINGTON] Harold Varmus, director of the National Institutes of Health (NIH), promised last week to review "in detail" the recommendations of a report saying that the agency should do more to involve the public in its research decision-making (see *Nature* 394, 111; 1998).

"We agree that there is room for additional public input" into how NIH sets research priorities, Varmus said in a statement. He added that the agency would review the recommendations of the report over the next few months and decide how to implement them. The report was produced by the Institute of Medicine at the direction of the US Congress.

Some disease-based organizations have given the report strong support. The Juvenile Diabetes Foundation International, for example, said it "may provide a foundation for regaining the public's confidence in the NIH priority setting process".

But David Korn, senior vice-president of the Association of American Medical Colleges, criticizes the proposal that a new council of public representatives should meet the NIH director regularly but not lobby him on research priorities. "As soon as you put people who feel strongly about such matters on such a council, it's exactly what they're going to try to do," says Korn.

...as Specter battles on for 14.7% funding boost

[WASHINGTON] The chairman of a key US Senate subcommittee has returned from cardiac bypass surgery to proclaim that he is still seeking a \$2 billion, 14.7 per cent increase for the National Institutes of Health (NIH) in the 1999 budget he is drawing up.

Senator Arlen Specter (Republican, Pennsylvania) chairs the subcommittee on Labor, Health and Human Services and Education of the Senate Appropriations Committee. He had earlier vowed to land such an increase for NIH. But this was before his subcommittee was told that it only had \$1.1 billion more to work with in funding a wider range of agencies than it did last year.

The corresponding House subcommittee has proposed a \$1.24 billion, 9.1 per cent NIH increase, but slashed several social programmes to achieve it.

Koplan to head Centers for Disease Control

[WASHINGTON] An epidemiologist who helped formulate the US response to the emergence of AIDS in the early 1980s was last week appointed director of the Centers

for Disease Control and Prevention (CDC), a \$2.3 billion agency with 6,900 employees.

Jeffrey Koplan, 53, is a former assistant surgeon general and the current president of the Prudential Center for Health Care Research in Atlanta. He will take up the new post on 5 October. He has long experience in public health, including chairing the Public Health Service's Executive Committee on AIDS from 1982 to 1984.

From 1989 to 1994, he was the first director of the CDC's National Center for Chronic Disease Prevention and Health Promotion. In that role, he established national early detection programmes for breast and cervical cancer.

French education budget escapes spending curbs

[PARIS] France's 'superministry', encompassing national education, research and technology, seems set to be a major winner in the state budget for 1999, scheduled to be announced on 22 July. According to the budgetary ceilings for each ministry, fixed last week, the education ministry would obtain a 4 per cent increase over its 1998 budget of FF374 billion (US\$61 billion). In contrast, the budgets of most ministries would be held constant as part of efforts to curb public spending.

The increase will be required for an ambitious proposal, currently under discussion, to refurbish and expand the universities. The proposal — *Universités du troisième millénaire* — would be launched next year, and be similar in scale to *Université 2000*, a scheme launched in 1991 under which FF32 billion was spent on building more than 1.5 million square metres of university space. The new plan would focus on universities in and around Paris, according to one official.

Indian nuclear experts get the cold shoulder

[NEW DELHI] The Indian government has admitted that a number of countries have imposed restrictions on visits by Indian nuclear scientists following the nuclear tests in May. The Canadian government, for example, has told the International Atomic Energy Agency in Vienna that it "would not welcome participation of nuclear experts from India in meetings in Canada until further notice".

Other countries are believed to have conveyed similar views informally or to have withheld visas. The United Kingdom has cut off all contact between its nuclear researchers and institutes and their counterparts in India and Pakistan. According to Govindarajan Padmanabhan, director of the Indian Institute of Science in Bangalore, "the supply of research samples from US government laboratories has been put on hold".

German universities to compete for extra cash

[MUNICH] Germany's research minister, Jürgen Rüttgers, last week promised to give universities an extra DM200 million (US\$109 million) in 1999. The money will be distributed on a competitive basis, and is primarily intended for multimedia equipment and for the establishment of courses compatible to Anglo-Saxon degrees.

There will also be more support for technology transfer schemes. The new grants will be controlled directly by the federal research ministry. The overall budget proposed by the cabinet would increase funding for the Department of Research, Education and Technology by 3.4 per cent.

Japan's Mighty Whale to harness tidal power

[TOKYO] A floating electricity plant known as the 'Mighty Whale', and designed to capture the energy of ocean tides, was launched in Japan last week. The floating plant is 50 metres long and 30 metres wide, and designed to use turbines to produce 110 kilowatts of electricity an hour. It was developed by the Science and Technology Agency and the Japan Marine Science and Technology Centre, and built by Ishikawajima-Harima Heavy Industry. It will be tested for two years in Gokasho Bay on Japan's southern coast.

UK museum completes £12m Earth Galleries



[LONDON] The opening this week of four new galleries in London's Natural History Museum marks the completion of a £12 million (US\$20 million) five-year redevelopment of the former Geological Museum into what are now known as the Earth Galleries.

The first gallery, 'From the beginning' (above), traces the geological evolution of the Earth from its origins in the 'big bang' 15 billion years ago. Exhibits include a meteorite studded with diamonds, and a rock sample returned from Mars. The galleries have been developed with a grant of £6 million from the Heritage Lottery Fund and £1 million from Rio Tinto Zinc plc.

When names are less than crystal clear

Sir — The problems of naming biological objects, such as genes, have been aired in *Nature* several times. Some of us, for example *Drosophila* geneticists, have been criticized for what has been seen as a “whimsical” attitude to gene names¹.

But there is now another problem. You published a Letter² by Kreusch *et al.* about the structure of the tetramerization domain of the *Shaker* potassium channel. *Shaker* is the “whimsical” name of a *Drosophila* gene that encodes a potassium channel, and was so called because mutant flies shake under etherization. It was, I believe, the first potassium-channel gene to be cloned.

In fields other than *Drosophila* genetics, the name *Shaker* is used to describe a family of potassium-channel proteins. All of your readers will (I hope) have several genes in their genome encoding proteins of this, and the related *Shab*, *Shaw* and *Shal* families. I have read every word of Kreusch *et al.*² in the hope of learning which species donated its gene for this (undoubtedly fine) study. Is it the *Drosophila melanogaster* *Shaker* protein? One of the human ones? Or perhaps from the lesser-eared bat, endemic to the lower caves on Mt Elgon in western Kenya? I have looked in the Protein Data Bank, only to find that the record is still being processed.

Do crystallographers have no concern for the variety of life? Have they no sympathy for those of us who are trying to build even bigger and better databases for biologists (I work on Flybase, a *Drosophila* database)? Are *Nature*'s word limits so

severe that a subeditor struck out the organism's name? Or do crystallographers simply not care? Whatever the answer, it is all very frustrating and, worse, bad practice.

Michael Ashburner

Department of Genetics, University of Cambridge, Cambridge CB2 3EH, UK

Senyon Choe and Paul Pfaffinger reply —

Our primary concern in reporting any scientific finding is to be accurate and precise. In a Letter to *Nature*, there is the additional requirement of brevity, while making findings accessible to a general scientific readership. Regarding the source of the *Shaker* protein, we erred on the side of accuracy, precision and brevity, but lost a general scientific reader, such as Ashburner. The protein source was identified at several points in our Letter¹ as the AK_v1.1a *Shaker*-type potassium channel, and a reference to the first use of this name in the literature was provided. A query of NCBI GenBank with the name AK_v1.1a correctly identifies the *Aplysia californica* channel sequence, and provides a page of further information.

A complete entry for our crystal structure was supplied to the Protein Data Bank, who were instructed to release this information upon publication. We believe that scientific advance thrives on openness, and we have an obligation to make available our material and data once published.

Finally, the use of the *Shaker* name to identify the channel subfamily, rather than a specific gene from *Drosophila*, is perhaps an

important topic for further discussion.

Because we entered into this problem from the viewpoint of trying to understand how the ordered assembly of potassium-channel subunits is restricted to hetero-multimerization with other proteins in the same subfamily, we see the importance of the *Shaker* subfamily name. Indeed, from our viewpoint it is remarkable that *Shaker* subfamily members selectively heteromultimerize across species, from *Aplysia* to human, and probably even elephants and lesser-eared bats (although the molecular biology of ion channels in the latter species has been woefully neglected).

We have produced a comparative analysis of the structural conservation of the T1 domain region for all 55 sequences that we could clearly identify in the gene database as *Shaker* members, as implicitly provided in Fig. 3b of the Letter². The complete alignment is accessible at: <http://sbl.salk.edu/~choe/clustal.html>. We hope to provide further information about this topic, and other implications of our research on the T1 domain, for both *Shaker* and non-*Shaker* channels, in future reports.

Senyon Choe

Structural Biology Laboratory, The Salk Institute, La Jolla, California 92037, USA

Paul Pfaffinger

Division of Neuroscience, Baylor College of Medicine, Houston, Texas 77030, USA

1. Jan, Y. N. *Nature* **389**, 665 (1997).

2. Kreusch, A., Pfaffinger, P. J., Stevens, C. F. & Choe, S. *Nature* **392**, 945–948 (1998).

Tokyo campus rising

Sir — While appreciating your interest in the University of Tokyo's plan for a new campus and the research departments that will occupy a central role in it, I would like to take issue with several comments in your coverage (*Nature* **392**, 429; & **393**, 5; 1998).

First, in response to your statements that we have had difficulty in obtaining the land for the Kashiwa campus, I would like to point out that this was partially acquired in 1995, and construction of the Institute of Solid State Physics is already under way. Land for the new research departments was purchased in April of this year through the government's supplementary budget.

Second, you argue that the commitment of funds to the new research departments has adversely affected the redevelopment of existing departments and institutes. Although there is a university-wide consensus on giving high priority to the new departments, we are also reviewing the priorities for the redevelopment of the

existing departments and institutes. Some of the redevelopment projects — such as the recently completed construction of the Faculty of Science Research Tower, the redevelopment of the Komaba campus, and the construction of a medical school building — are already becoming a reality.

Finally, you comment on the question of age limits for faculty members recruited to the new research departments. Although it was indicated that promising young scholars were being sought for certain associate professor positions, such an age limit is not a requirement in all fields.

Every effort is being made to ensure that the successful setting up of the departments is a goal of the entire university community, and we believe firmly that they represent a welcome change from the present rigid structure of graduate education and research in Japan. The new research departments were inaugurated in April.

Shun-ichi Kobayashi (Vice-President)

University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8654, Japan

Haldane's speculation

Sir — The retrospective review, by David Jones, of *The Scientist Speculates* says “how they kept J. B. S. Haldane out of it [the book] remains a mystery” (*Nature* **393**, 642; 1998). I am writing to resolve that mystery.

A nice article by Haldane was accepted, but he was annoyed when the publishers put out an advance leaflet in which Arthur Koestler was put at the head of a list of scientific contributors. Haldane regarded that as inappropriate and withdrew his article. It was published later in *Science News* (Penguin).

On another point, my guess in 1962 for the cost of reaching full artificial intelligence by 1978 was \$10^{8.7 ± 1}, so the 95% interval was from \$5 million to \$50 billion. This guess was not refuted!

I. J. Good

Department of Statistics, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061-0439, USA

How to make microbes safer

Much could be gained by building closer links between international efforts to protect society from natural and engineered microorganisms, and parallel efforts to prevent the threat of microorganisms as biological weapons.

Graham S. Pearson

It is widely accepted that the potential applications of microorganisms through biotechnology offer immense benefits in medicine, food production and other fields. But public concern is nevertheless, and rightly, intense. On the one hand, there is a general awareness of the dangers of disease, for example Ebola virus outbreaks in Zaire or bovine spongiform encephalopathy in cattle and its possible spread to humans; and there is equally broad, and growing, concern about the consequences of genetic engineering, for example in modified foodstuffs. In parallel, there is also concern that disease may be spread deliberately, for example in war or through a terrorist attack.

In all such instances, public concern stems from the way in which microorganisms are stored, transferred and used. Many countries in the developing world in particular are keen to benefit from advances in microbiology, yet wish to protect their people and environment from harmful pathogens. Similarly, industrialized countries want to ensure that biological weapons are prohibited and so could not be an option either for governments or for terrorist groups.

In both cases, the question we face is how the undoubted benefits of biotechnology be obtained without risk. One development that would, I feel, help to address this dilemma would be to explore ways in which current efforts to achieve the two objectives described in the previous paragraph can be brought closer together.

Health and environment

International efforts to regulate the potential environmental damage of manmade microorganisms have centred around agreement, reached at the Earth Summit in Rio de Janeiro in 1992, on a set of principles intended to achieve sustainable development while protecting the environment. These principles include: increasing the availability of food, feed and renewable raw materials; improving human health; enhancing the protection of the environment; establishing enabling mechanisms for the development and environmentally sound application of biotechnology; and enhancing safety and developing international mechanisms for cooperation.

The last of these principles explicitly requires the further development of internationally agreed principles on risk assessment and management of all aspects of biotechnology. The document agreed in Rio says



Battling bugs: US troops simulate tests for the presence of biological warfare agents such as anthrax.

that "only when adequate and transparent safety and border-control procedures are in place will the community at large be able to derive maximum benefit from, and be in a much better position to accept the potential benefits and risks of, biotechnology".

Safety protocol

The Convention on Biological Diversity opened for signature at the Rio summit currently has 174 ratifications (a notable exception being the United States). One article of the convention addresses the international distribution of the benefits of biotechnology, and includes consideration of its safety and transfer aspects. It states that: "The parties shall consider the need for and the modalities of a protocol setting out appropriate procedures, including, in particular, advance informed agreement, in the field of the safe transfer, handling and use of any living modified organism resulting from biotechnology which may have adverse effect on the conservation and sustainable use of biological diversity."

The biodiversity convention is implemented by conferences of these parties. At their first meeting, held in Nassau in the Bahamas in 1994, a group of specialists on biosafety appointed by the parties was asked to consider how to produce a protocol of appropriate procedures. Such a protocol would need to cover advance informed agreement in the field of safe transfer, han-

dling and use of any living modified organism resulting from biotechnology that may have an adverse effect on conservation, and the sustainable use of biological diversity.

A two-track approach was agreed at the biodiversity conference at its second meeting, held in Djakarta, Indonesia, in 1995. A working group was set up to begin negotiations on a biosafety protocol, aiming for completion by 1998. In parallel, efforts would be made to finalize the International Technical Guidelines on Safety on Biotechnology of the United Nations Environment Programme (UNEP).

These two initiatives are intended to be complementary. The protocol will be a binding agreement setting out what countries must do, while the UNEP guidelines are intended to indicate how to do it. The biosafety working group will have a protocol ready for adoption by the convention signatories after the final biosafety meeting in February 1999 in Montreal. UNEP's international guidelines were adopted by a meeting of government-designated experts in Cairo in 1995, and subsequently issued by UNEP. They have since been developed at further meetings.

Security initiatives

International efforts to control the potential military use of microorganisms have taken place on a separate track. Biological warfare — the use of disease as a weapon against

humans, animals and plants — was the first class of weapons to be totally banned under the Biological and Toxin Weapons Convention of 1972. The scope of this convention is very broad, and currently there are 140 'states parties' and 18 signatories; the United Kingdom, United States and now Russia are the co-depositaries of the convention. Iraq ratified the convention in 1991 after the Gulf War. Signatory states undertake not to develop, produce, stockpile or otherwise acquire or retain microbial or other biological agents or toxins, whatever their origin or method of production, "of types and in quantities that have no justification for prophylactic, protective or other peaceful purposes".

Successive review conferences of the biological weapons convention have reaffirmed that the ban extends to "all relevant scientific and technological developments, *inter alia*, in the fields of microbiology, biotechnology, molecular biology, genetic engineering and any applications resulting from genome studies, and the possibilities of their use for purposes inconsistent with the objectives and the provisions of the convention".

But the biological weapons convention lacks teeth: it cannot monitor and verify compliance, and so cannot enforce its own rules. Although politically binding, confidence-building measures were agreed in 1986 and extended in 1991, only just over half the states parties have made even one annual declaration of information to be provided on specified, relevant activities, and only about a dozen states have made the agreed annual declarations — and these vary significantly in quality.

At the most recent review conference of the convention in December 1996, the United States expressed concern over its assessment that "twice as many countries now have or are actively pursuing offensive biological weapons capabilities as when the convention went into force". Indeed, it is clear that, despite the existence of the convention, biological weapons continue to be attractive to some states. In 1992, for example, Russia admitted that the former Soviet Union, despite being a co-depositary of the biological weapons convention, had continued what was later described by a UK Foreign Office minister as "a massive offensive biological weapons programme conducted illegally for years".

The UN Special Commission on Iraq has, through determined efforts, uncovered at least part of the sizeable and significant biological weapons programme which president Saddam Hussein has tried to keep hidden. In its report in October 1997 to the UN Security Council, the commission said that the Iraqi full, final and complete disclosure on such weapons delivered the previous month "fails to give a remotely credible account of Iraq's biological warfare programme".

Of all weapons of mass destruction, bio-

"Twice as many countries now have or are pursuing offensive biological weapons capabilities as when the convention went into force".

logical weapons today present the greatest danger. The mechanism to prevent them is weak and the weapons are extremely easy to acquire, yet their effects would be comparable to those of nuclear weapons. It was an awareness of this issue that led the 1991 review conference to set up a group of government experts to examine the scientific and technical aspects of possible verification measures.

The 1993 report of this group (called VEREX) led to the creation of a group to draft verification proposals to strengthen the convention, which would be included "as appropriate" in a legally binding instrument. This group has since produced a "rolling text" of a possible draft protocol, currently comprising 23 articles and 8 annexes. The draft contains all the elements for an effective regime, including mandatory declarations, 'non-challenge' visits (random and focused) of declared facilities, and compliance concern investigations (facility and field).

In addition, the group's mandate requires consideration of measures to ensure full implementation of an article of the convention in which states undertake to facilitate — and have the right to participate in — the fullest possible exchange of equipment, materials, and scientific and technological information for the use of bacteriological agents and toxins for peaceful purposes.

Various changes to the text of the verification proposals were suggested by some industrialized countries such as the United States at the most recent meeting in March this year. These changes cast doubt unnecessarily on the group's intention to address the problem of how to promote the peaceful use of microorganisms yet build confidence that such materials are not being used for prohibited purposes, and have been severely criticized by many developing countries.

Regulatory frameworks

The potential threat of dangerous diseases to people, livestock and crops was, of course, recognized long before the 1992 Rio summit. But Rio provided a welcome focus on these issues and, as a result, many countries have introduced national regulations to control the handling, use, storage and transfer of hazardous pathogens. More recently, various countries have tried to harmonize their national regulations both regionally and

internationally to facilitate cooperation and trade.

Under such regulations, those wishing to work with pathogenic organisms are usually required to provide information to national authorities, which carry out inspections of the facilities in which the pathogens are to be handled. Such notification and inspection can be required before any work is carried out. In addition, transfers of potentially dangerous pathogens are controlled and monitored in several countries, while many others have additional regulations on the handling and use of genetically modified organisms.

There are parallel regulations for licensing facilities engaged in producing medical products for humans and veterinary products for animals, to ensure that the products are safe and effective. These require repeated inspection of facilities to ensure that standards of good manufacturing practice are met. Such regulations should promote trade in pharmaceuticals, and in building public confidence that such facilities are not being used for prohibited purposes.

The net effect of all these controls and regulations is to provide national frameworks intended to ensure that the danger presented by pathogens to public health and the environment is minimized, and that medical products are safe, so providing assurance to the public. There are various moves under way to harmonize such measures both regionally and internationally, as an outbreak of human, animal or plant disease in one country can easily spread to others.

Objectives

One objective of international security is to ensure that the use of disease as a weapon of war is prohibited, and to devise a regime to monitor compliance with this ban, to detect and deter offenders. The pathogens that can be used as weapons are precisely those that present a danger in natural outbreaks of disease. Any regime to strengthen the biological weapons convention should have two important elements. One is that declarations about facilities and activities of particular relevance to the convention would be made to national authorities and to the international organization responsible for enforcement. The second is that these declarations should be confirmed by occasional visits.

There are common goals between the current negotiations to develop protocols on the 'peaceful' use of microorganisms, in Montreal, and on their 'prohibited military' use, in Geneva. There are synergistic benefits to be won, for health and environmental safety and for international security, by strengthening the biological weapons convention. □

Graham S. Pearson is in the Department of Peace Studies, University of Bradford, Bradford BD7 1DP, UK. He was formerly director general and chief executive of the UK Chemical and Biological Defence Establishment, Porton Down.

Dust hides a universal starburst

Douglas Scott

When were the stars born? A new type of detector has seen powerful emission from dusty distant galaxies, indicating that the most fertile period of star formation occurred when the Universe was only about an eighth of its current age.

Modern cosmology asks two questions. What sort of Universe do we live in, in terms of size and shape and constitution? And how did galaxies and larger structures form in it? There is great optimism that we may soon know a lot about the first question — the density of the Universe and how fast it is expanding, for example. The details of structure formation, on the other hand, are likely to remain murky for a long time. But one piece of this puzzle may now be getting clearer.

Astrophysics has flourished this century, as new types of detector have opened up more and more of the electromagnetic spectrum. By exploring different wavelength bands, different ranges of physical conditions are probed. Dust is one example. Grains of silicon, carbon and other interstellar detritus, up to a few hundred micrometres across, absorb visible and ultraviolet light. That warms them up to a few tens of kelvin, and so they re-emit the energy in the far-infrared band (Fig. 1).

A new camera, designed to operate in this relatively unexplored region of the electromagnetic spectrum, has revealed a surprising fact: dust may be the most luminous stuff in the Universe. Papers on pages 241 and 248 of this issue, by Hughes *et al.*¹ and by Barger *et al.*², describe what appear to be dusty young galaxies existing at a time when the Universe was about an eighth of its present age. That means that normal galaxies were forming the bulk of their stars much earlier than we thought. This early, rapid star formation implies that the very first objects in the Universe might have had to form much earlier still, which could put a tight constraint on theories of galaxy formation.

We expected young galaxies to be dusty, and therefore, if they were also forming stars at a high rate, to be infrared-bright objects. The expansion of the Universe would shift that emission into the submillimetre part of the far-infrared band. (By the time it reaches us, light from an object at a redshift of z has been stretched by a factor of $(1+z)$ by the expansion of space. So light with a wavelength of $200\text{ }\mu\text{m}$ emitted by a galaxy at $z=3$ appears to us as submillimetre radiation, at $800\text{ }\mu\text{m}$.)

The Cosmic Background Explorer (COBE)

was the first instrument in this waveband sensitive enough to detect any distant emission. Signs of a powerful far-infrared background (FIB) were first discovered in COBE data³ in 1996. The strength of this background⁴ indicated that, at some time in the early Universe, dust was emitting more radiation than all the visible stars. The FIB can only come from dusty, star-forming galaxies distributed throughout space.

But at what epoch was most of this

emission produced? Optical spectra of single objects are needed to determine redshift precisely (matching observed spectra with known patterns of emission and absorption lines). But to pick out individual galaxies would require an angular resolution higher than COBE's. While such an instrument was being developed, another route was being taken to determine the era of star formation.

Astronomers had been developing ways of using other wavelengths to find high-redshift galaxies. One technique is to select objects that have disappeared in the near ultraviolet, because all their near-ultraviolet emission has been redshifted into the optical or infrared wavebands. This 'ultraviolet-dropout' method has been used⁵ to find many galaxies at redshifts of about 3. From their brightness and number density, it appeared that the star-formation rate of the Universe peaked at $z \approx 1.5$, when the Universe was about half its present age. But to make this estimate, a highly uncertain correction for dust obscuration has been made. If the high-redshift galaxies contain more dust

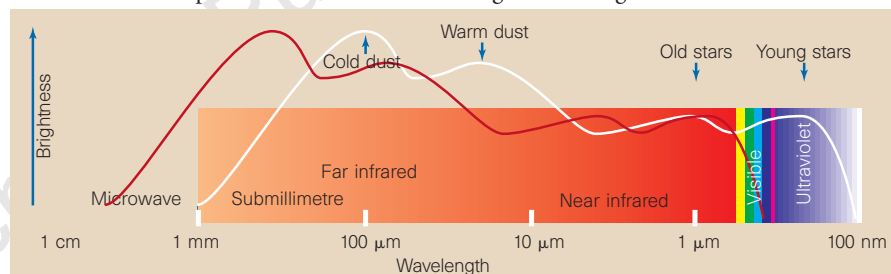


Figure 1 The electromagnetic spectrum, from microwave to near ultraviolet. The curve is a schematic spectrum of a rapidly star-forming galaxy, such as Arp220 (Fig. 3, overleaf). At redshifts of around 3, the peak of the cold dust emission is moved well into the submillimetre band, and the emission from young stars is shifted out of the ultraviolet.

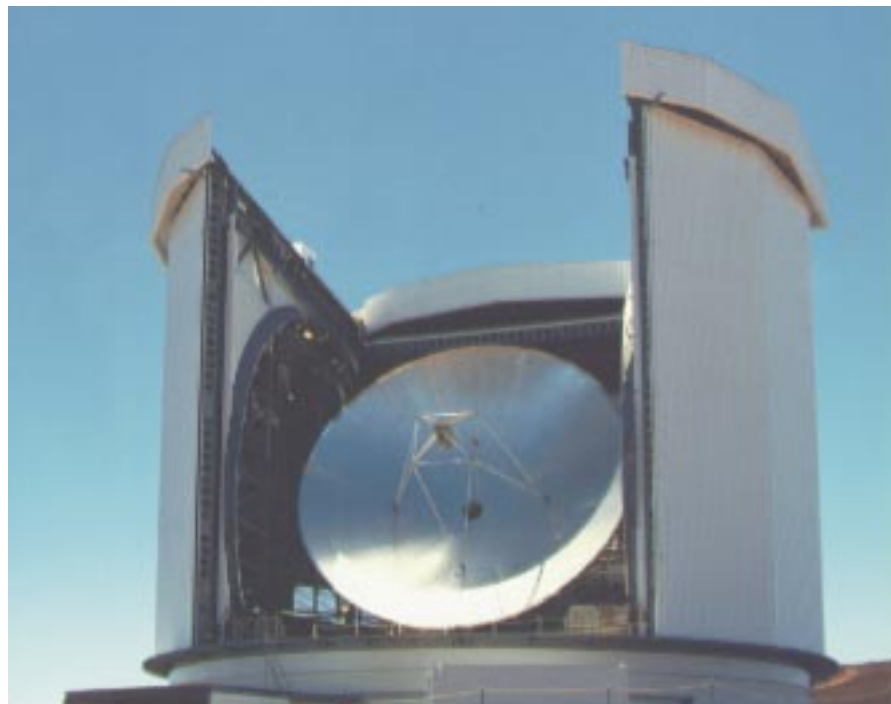


Figure 2 Short-wave receiver: the James Clerk Maxwell Telescope in Hawaii, which collects submillimetre radiation from space, notably that emitted by dust at temperatures of a few tens of kelvin.

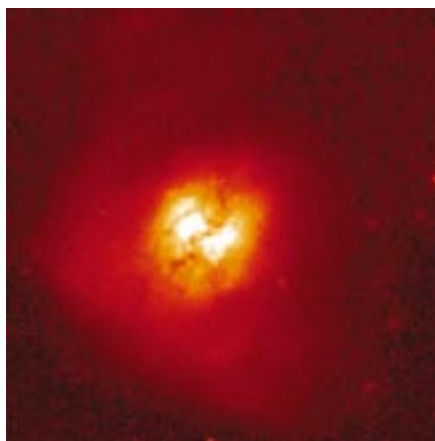


Figure 3 The starburst galaxy Arp220. Dust traps much of its starlight, and re-radiates the energy in the infrared. In the early Universe, most galaxies may have been like this.

than expected, then the overall star-formation rate may have been drastically underestimated, and the real peak may lie further in the past. So can we measure the effect of this dust obscuration by observing the dust emission directly?

The Submillimetre Common-User Bolometer Array⁶ (SCUBA) was installed on the James Clerk Maxwell Telescope (Fig. 2) in late 1996. It came along at the perfect time to bring together the FIB and optical studies. SCUBA is a large array of bolometers (detectors that measure the heat deposited in them) operating at an extremely low temperature, below 100 mK. The number of detectors, and the high sensitivity gained by being so cool, enable SCUBA to collect data thousands of times faster than its predecessors. This instrument works at a range of wavelengths, but is optimized for 850 μm , perfect for detecting dust emission from high-redshift galaxies.

In all, Hughes *et al.*¹ and Barger *et al.*² report seven objects detected by SCUBA, confirming earlier predictions⁷ that there would be several thousand 'SCUBA-bright' objects per square degree of sky. Overall, their far-infrared emission outshines the optical emission of ultraviolet-dropout galaxies by at least a factor of five, implying that dust emission dominates total light output at these redshifts.

Collectively, these dust-emitting galaxies would make up about half of the background measured by COBE, and the rest is probably contributed by sources that are a little fainter. We still don't know whether there is any overlap between this population of SCUBA objects and those selected in the optical waveband as ultraviolet dropouts. But the FIB has now been broken up into its constituent galaxies, and the peak of star formation in the Universe now appears to be near a redshift of 3.

The angular resolution of the James Clerk Maxwell Telescope is much cruder than that

of an optical telescope, and so it is difficult to identify optical counterparts to the galaxies seen by SCUBA (Fig. 3). So no definitive redshift has been measured for any of these seven new objects; some of them may be at even higher redshifts, pushing the main time of star formation still further back.

This question should be resolved by a new generation of submillimetre telescopes with higher angular resolution and collecting area, or by space-based missions that can probe a wider range of wavelengths. Several such experiments are expected soon, building towards the joint Large Southern Array/Millimetre Array project, and the Far-Infrared and Submillimetre Telescope satellite, both expected in about ten years. Meanwhile, further optical and infrared observations will be needed to show whether smaller

objects formed first before merging into large galaxies, and what part dark matter had to play. Through a combination of future submillimetre and optical studies, we may yet learn when and how the building blocks of the Universe were assembled. □

Douglas Scott is in the Department of Physics and Astronomy, University of British Columbia, 6224 Agricultural Road, Vancouver, British Columbia V6T 1Z1, Canada.

e-mail: dscott@astro.ubc.ca

1. Hughes, D. H. *et al.* *Nature* **394**, 241–247 (1998).
2. Barger, A. J. *et al.* *Nature* **394**, 248–251 (1998).
3. Puget, J.-L. *et al.* *Astron. Astrophys.* **308**, L5–L8 (1996).
4. Fixsen, D. J., Dwek, E., Mather, J. C., Bennett, C. L. & Shafer, R. A. *Astrophys. J.* (in the press); preprint astro-ph/9803021 on xxx.lanl.gov.
5. Steidel, C. C., Giavalisco, M., Pettini, M., Dickinson, M. & Adelberger, K. L. *Astrophys. J.* **462**, L17–L21 (1996).
6. <http://www.jach.hawaii.edu/JCMT/scuba/>
7. Smail, I., Ivison, R. J. & Blain, A. W. *Astrophys. J.* **490**, L5–L8 (1997).

Signal transduction

Why Ras needs Rho

Frank McCormick

Members of the Rho family of GTPase proteins regulate an assortment of processes: they rearrange the cytoskeleton; they activate lipid kinases, protein kinases and phospholipase D; and they can regulate transcription (reviewed in ref. 1). They are also essential for transformation of cells to an oncogenic state by another GTPase, Ras², although it is not known which of Rho's many functions Ras actually needs. On page 295 of this issue, however, Olson, Paterson and Marshall³ describe a new effect of Rho, which explains how Rho helps Ras to transform cells.

Rho proteins are best known for their ability to bundle actin into stress fibres, and promote the assembly of focal adhesions where these fibres connect with cell-surface integrins. But these are clearly not the contribution that Rho makes to oncogenic transformation — cells transformed by Ras usually lack both stress fibres and focal adhesions^{1,4}. In these cells, Rho must be doing something else. Olson *et al.*³ now show that Rho suppresses p21^{Waf1/Cip1}, a well-known inhibitor of the cell cycle, thus allowing Ras to drive cells into S phase. What's more, they show that this is the only function of Rho that really matters in the context of Ras-induced DNA synthesis.

These provocative results raise several questions. First, how is the activity of Rho controlled under physiological conditions? This is important because the data indicate that the biological effects of Ras will depend, to some degree, on the amount of active Rho in a cell. In cultured fibroblasts, a major pathway for turning Rho on comes from lysophosphatidic acid (LPA), which itself potently stimulates cell division. LPA acts through a guanine-nucleotide-binding (G)-protein-coupled receptor¹, although the

pathway to Rho is unknown. But, because LPA (along with various growth factors and nutrients) is a component of serum, we might expect that transformation of cells by Ras depends on serum. In fact, the degree to which different Ras-transformed cells depend on serum varies immensely, and this may be because other pathways lead to activation of Rho. One of them could come from Ras itself, for when oncogenic Ras is injected into cells Rho becomes active (stress fibres and focal adhesions assemble)⁵. However, this effect is transient and the extent to which the Ras–Rho pathway is used in stable, transformed cells is not known.

The second question is, how does Rho prevent induction of p21^{Waf1/Cip1}? Olson *et al.* show that, to some extent, the effects of Rho are exerted on transcription of p21^{Waf1/Cip1}.

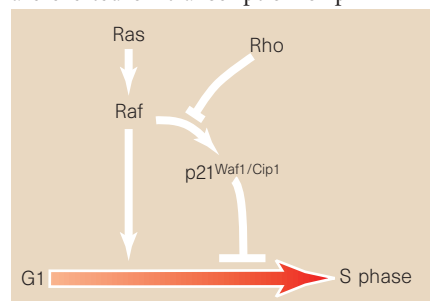


Figure 1 Regulation of Ras-dependent DNA synthesis by Rho. Ras activates the serine/threonine kinase Raf, sending positive signals down a kinase cascade (the mitogen-activated protein kinase pathway) and prompting cells to move through G1 of the cell cycle towards S phase. Excessive signalling from Ras/Raf induces p21^{Waf1/Cip1} and blocks entry into S phase. Olson *et al.*³ have now found that Rho overcomes the cell-cycle block by suppressing expression of p21^{Waf1/Cip1}.

Unfortunately, this does not offer a mechanistic explanation. Rho has already been implicated in transcriptional regulation, as a step in signalling from the LPA receptor to the serum response factor, which binds to the serum response element of the *fos* promoter⁶. But we do not know how Rho transmits the signal, or how these signals are received. Furthermore, Olson *et al.* leave open the possibility that Rho regulates levels of p21^{Waf1/Cip1} at the post-translational level as well.

To understand how Rho interferes with induction of p21^{Waf1/Cip1}, we first need to ask how Ras itself mediates this effect. This has already been answered in part. Ras binds and activates the serine/threonine kinase Raf, which sends signals down a cascade of kinases that leads, eventually, to DNA synthesis. However, excessive Raf activity induces p21^{Waf1/Cip1} and so blocks further DNA synthesis⁷ (Fig. 1). This feedback effect must limit the ability of Ras or Raf to transform normal cells (because transformation requires DNA synthesis). Consistent with this, activated Rho gives a tremendous boost to Raf's transforming power², and its effects on p21^{Waf1/Cip1} could now explain this. Unfortunately, this explanation falls short at the molecular level because we do not know how excessive signalling down the kinase cascade induces p21^{Waf1/Cip1}.

At present, the best-known regulator of p21^{Waf1/Cip1} is p53. This protein activates p21^{Waf1/Cip1} when cells are exposed to DNA-damaging agents, allowing the cells time to fix the damage before they commit to S phase. Unfortunately, Raf-dependent induction of p21^{Waf1/Cip1} does not require p53, at least in mouse embryonic fibroblasts⁷, and the precise mechanism of induction remains obscure.

So far, Olson and co-workers have focused on how Rho helps Ras to drive cells into S phase. But oncogenic transformation by Ras clearly depends on many other physiological events, including those driven by other Ras effectors. For example, the ability of transformed cells to survive detachment and grow as colonies in soft agar depends on activation of phosphatidylinositol-3-OH kinase by Ras⁸. Is Rho involved in this, or other aspects of oncogenic transformation? Rho is not needed for Ras-driven DNA synthesis in cells that lack p21^{Waf1/Cip1}, so is Rho dispensable for Ras-driven transformation in these cells as well? If so, we would have to conclude that Rho's only function in oncogenic transformation is to suppress p21^{Waf1/Cip1} — unlikely, but possible. In any event, the pathways that connect Rho to the expression of p21^{Waf1/Cip1} will now be studied. We need to know which of Rho's many effectors initiates this pathway. Analysis of Rho proteins with mutated effector binding sites points towards the protein kinase p160ROCK as a candidate⁹, but where this pathway leads is anyone's guess. □

Frank McCormick is at the Cancer Center, University of California at San Francisco, 2340 Sutter Street, San Francisco, California 94115, USA. e-mail: mccormick@cc.ucsf.edu

1. Van Aelst, L. & D'Souza-Schorey, C. *Genes Dev.* **11**, 2295–2322 (1997).
2. Qiu, R. G., Chen, J., McCormick, F. & Symons, M. *Proc. Natl Acad. Sci. USA* **92**, 11781–11785 (1995).

3. Olson, M. F., Paterson, H. F. & Marshall, C. J. *Nature* **394**, 295–299 (1998).
4. Hughes, P. E. *et al. Cell* **88**, 521–530 (1997).
5. Ridley, A. & Hall, A. *Cell* **70**, 389–399 (1995).
6. Hill, C. S., Wynne, J. & Treisman, R. *Cell* **81**, 1159–1170 (1995).
7. Woods, D. *et al. Mol. Cell. Biol.* **17**, 5598–5611 (1997).
8. Khwaja, A. *et al. EMBO J.* **16**, 2783–2793 (1997).
9. Sahai, E., Alberts, A. S. & Treisman, R. *EMBO J.* **17**, 1350–1361 (1998).

Planetary impacts

Craters unchained

H. J. Melosh

Ever since the appearance of those beautiful 'string of pearls' images, created by the breakup of comet Shoemaker-Levy 9 in July 1994, planetary geologists have been excitedly looking for chains of impact craters caused by similar events. A good case can indeed be made that the lines of craters called 'catenae' on the Jovian satellites Ganymede and Callisto were created by the impact of other tidally split comets¹. There are even two crater chains on Earth's moon that may have been caused by an asteroid or comet originally split up by the Earth's tidal forces^{2,3}.

But what of crater chains on the Earth itself? Several candidate alignments of terrestrial craters have been proposed, and these putative relatives of catenae have been eagerly pursued by both science journalists and planetary scientists. However, a critical appraisal of one of the most extensive proposed crater chains on Earth, published by John Luczaj in *Geology*⁴, will tend to head off the stampede towards interpreting aligned dents in the Earth as the result of impacts.

Shortly after the recognition that catenae on Ganymede and Callisto are the likely result of collisions with tidally split comets, Rampino and Volk⁵ identified a possible alignment of eight crater-like structures stretching 700 kilometres across the states of Kansas and Missouri, and into southern Illinois (Fig. 1, overleaf). Two of these structures, Decaturville (3 in Fig. 1) and Crooked Creek (5), had previously been identified as the result of impacts that occurred some 300 million years ago, but the origins of the others were less clear. Luczaj⁴ has searched the literature on each of the other structures and derived constraints on their ages. The members of a true impact crater chain must have identical geological ages because they form within minutes of one another. Luczaj found, however, that the ages of the structures in Rampino and Volk's string could not be the same and that an impact origin is therefore impossible. Instead, he argues that these craters are of volcanic origin and that the alignment is purely tectonic, controlled by deep fractures in the Earth's crust.

Luczaj's result may disappoint many crater-chain enthusiasts. But it offers wel-

come relief to theoreticians, who have been hard-pressed to explain all of the crater chains that have been proposed lately. For example, Ocampo and Pope⁶ suggested an impact origin for three aligned circular structures imaged by the Space Shuttle radar in northern Chad; a few months ago, Spray *et al.*⁷ announced an alignment of the palaeolatitudes of several large impact craters, including the 90-km-diameter Manicouagan in Quebec and Rochechouart in France; and there is also a nearly forgotten alignment of the craters Auouelloul, Tenoumer and Temimichat Ghallaman in Mauritania⁸.

The problem with these putative chains of impact craters is that theoreticians have no idea how they could have formed. So if they are indeed impact chains, they may be pointing to a serious deficiency in our understanding of impact processes. The burden of proof is on the proponents of such chains, however, as we must be very sure that they really are the products of a series of near-simultaneous impacts.

Impact crater chains form only under very special conditions. The fact that they are clustered in both time and space means their parent body must have broken up just a short time before the collision. If the parent body breaks up too far away from the target planet, only one or two fragments will ever strike it, most missing their target and flying by into space. This is the case for comets that break up at solar perihelion — their fragments are too dispersed to make crater chains by the time they reach any planetary body⁹.

On the other hand, if breakup occurs too close to impact, the fragments will not disperse enough to form distinct craters. This is the case for large incoming meteorites affected by the Earth's own tides or aerodynamic forces during entry¹⁰. Small meteorites, less than about 100 m in diameter, do form characteristic strewn fields of aerodynamically dispersed fragments, but these are easy to distinguish from crater chains¹¹. Likewise, doublet craters are relatively common on Earth, but appear to have been created by binary parent bodies whose components are gravitationally bound together before impact¹².

Furthermore, most breakup processes



Figure 1 Geological structures in the midwestern United States that can be attributed to impacts (stars) or to probable volcanic activity (circles). The Rampino–Volk crater chain re-analysed by Luczaj⁴ is numbered 1–8, with Decaturville being 3 and Crooked Creek 5. A point not discussed in the main text is that, curiously, the southernmost extent of Pleistocene glacial deposits (green line) parallels the Rampino–Volk chain for its entire length from Kansas to southern Illinois, and passes over three other structures running from North Dakota to Ohio (A–C). Whether these coincidences are significant or are just chance remains a completely open question. It is, however, much easier to discover bedrock structures in old rocks not covered by glacial drift, which may account for the large density of both impacts and volcanic structures south of the terminal moraine. Another questionable crater chain could be claimed for four structures in upper Michigan and Wisconsin (i–iv), which is also curiously close to the ‘driftless region’ in Wisconsin. (Data from refs 20–22.)

generally produce clusters of craters, not linear chains. Tidal forces, exerted strictly radially to the perturbing planet, seem almost uniquely capable of pulling fragments into long chains. Thus, a dive of a comet or asteroid close to Jupiter, linear tidal stretching of the fragments¹³ and an immediate splash onto the surface of a nearby large satellite has produced several dozen catenae on Ganymede and Callisto over the course of Solar System history¹⁴.

A similar close pass to Earth and a splash onto our relatively large moon may have produced one or two catenae on the Moon’s near side. A reciprocal close pass to the Moon and impact on Earth might have occurred once in all of Solar System history¹⁵, but that is obviously inadequate to produce all of the three or four crater chains proposed to have been created in the past 500 million years. Very straight crater chains are occasionally formed as secondaries to large impact events², but the existence of a fresh chain of secondary craters requires the nearby presence of a much larger primary of the same age, something that is lacking for all of the proposed terrestrial crater chains. Even more bizarre possibilities — such as a dip of the impactor through Earth’s atmosphere, with enough loss of momentum to enable its capture — have also been considered, but these also seem highly improbable¹⁶. In this scheme, the impacts would all have to be at

very low angles, which is not apparent for any of the proposed chains.

Under these circumstances, Luczaj’s analysis⁴ is valuable — in possibly discrediting the reality of the Rampino and Volk impact chain, it may relieve us of the need to explain it by impact processes. Luczaj has, however, gone a bit too far in casting doubt on the impact origins of every member of the chain. The impact origin of two members, Decaturville and Crooked Creek, has been accepted by experts for many years, whereas doubt about the other structures has been widespread. Luczaj did not visit any of the structures himself, nor did he undertake any new petrographic analysis of samples, but relied instead on published descriptions.

He also considers only two hypotheses about the origin of the proposed chain — that they are either all volcanic or all impact craters — seemingly dismissing the possibility that they could be of mixed origins. This assumption of a common origin for all members of the chain logically leads Luczaj to question the currently accepted criteria for recognition of an impact and to raise the old spectre that volcanic eruptions might generate shock features, which are taken to be diagnostic of impact craters. This is a view that has been repeatedly raised and just as repeatedly laid to rest by planetary geologists for years¹⁷.

Although the probability of a coincidence between two bona fide impact craters and a

pre-existing line of tectonic structures may seem low, it appears that just such a coincidence must have occurred. This type of argument has been used before to ‘disprove’ the impact origin of craters — in 1953, Dorsey Hager¹⁸ argued that Meteor Crater, Arizona, is of volcanic origin on the basis of its coincidence with a pre-existing anticline in the bedrock. But as the late crater expert Gene Shoemaker was fond of pointing out, nearly any randomly chosen point is ‘special’ in some geological sense, making this an invalid argument for, or against, the impact origin of any given structure. The true probability of a given association or alignment has to be carefully evaluated against the possibility of systematic biases in occurrence, exposure or recognition of the structures in question.

And what of Spray and colleagues’ crater alignment⁷? This does at least involve verified impact structures. But to form a physically associated crater chain the component craters must be of identical geological age. Based on new radiometric ages, Spray *et al.* argued that this is the case. Dennis V. Kent, however, has data which he believes show that the two major craters in the chain, Manicouagan and Rochechouart, may have formed in eras of opposite magnetic polarity — which would thus cast doubt on this chain also.

As for the others, only careful geological investigation will reveal whether impact craters really do occasionally form aligned chains on the Earth, or whether the theoreticians are right after all, and that such associations are too improbable to exist. □

H. J. Melosh is in the Lunar and Planetary Laboratory, University of Arizona, Tucson, Arizona 85721, USA.

e-mail: jmelosh@lpl.arizona.edu

- Melosh, H. J. & Schenk, P. *Nature* **365**, 731–733 (1993).
- Melosh, H. J. & Whitaker, E. A. *Nature* **369**, 713–714 (1994).
- Wichman, R. W. & Wood, C. A. *Geophys. Res. Lett.* **22**, 583–586 (1995).
- Luczaj, J. *Geology* **26**, 295–298 (1998).
- Rampino, M. R. & Volk, T. *Geophys. Res. Lett.* **23**, 49–52 (1996).
- Ocampo, A. C. & Pope, K. O. *Lunar Planet. Sci.* **XXVII**, 977–978 (1996).
- Spray, J. G., Kelley, S. P. & Rowley, D. B. *Nature* **392**, 171–173 (1998).
- Fudali, R. F. & Cressy, P. J. *Earth Planet. Sci. Lett.* **30**, 262–268 (1976).
- Hut, P., Shoemaker, E. M., Alvarez, W. & Montanari, A. *Proc. Lunar Planet. Sci. Conf.* **XXII**, 603–604 (1991).
- Melosh, H. J. & Stansberry, J. A. *Icarus* **94**, 171–179 (1991).
- Passy, Q. & Melosh, H. J. *Icarus* **42**, 211–233 (1980).
- Bottke, W. F. & Melosh, H. J. *Nature* **381**, 51–53 (1997).
- Asphaug, E. & Benz, W. *Nature* **370**, 120–124 (1994).
- Schenk, P. M., Asphaug, E., McKinnon, W. B., Melosh, H. J. & Weissman, P. R. *Icarus* **121**, 249–274 (1996).
- Bottke, W. F., Richardson, D. C. & Love, S. G. *Icarus* **126**, 470–474 (1997).
- Love, S. G., Bottke, W. F. & Richardson, D. C. *Lunar Planet. Sci.* **XXVIII**, 837–838 (1997).
- Grieve, R. A. F., Langenhorst, F. & Stöffler, D. *Meteoritics Planet. Sci.* **31**, 6–35 (1996).
- Hager, D. *Bull. Am. Assoc. Petrol. Geol.* **37**, 821–857 (1953).
- Kent, D. V. *Nature* (Scientific Correspondence, in the press).
- Grieve, R. A. F. & Shoemaker, E. M. in *Hazards due to Comets and Asteroids* (ed. Gehrels, T.) 417–462 (Univ. Arizona Press, Tucson, AZ, 1994).
- Memmi, J. M. *GIS World* **8**, 54–57 (1995).
- King, P. B. & Beikman, H. M. *1:2,500,000 Geologic Map of the United States* (US Geol. Surv., Reston, VA, 1974).



100 YEARS AGO

The current number of the *Lancet* has a note interesting to the vast army of cyclists. After a "spin" along a more or less dusty road the cyclist sometimes experiences a dry and subsequently sore and inflamed throat. Headache and depression often follow, and the symptoms generally simulate poisoning of some kind. When the bacteriology of road dust is considered, these effects are hardly to be wondered at. ... Indeed, there can be no reason for doubting the infective power of dust when it is known that amongst the microbes encountered in it are the microbes of pus, malignant oedema, tetanus, tubercle, and septicæmia. The mischief to riders as well as to pedestrians would probably be largely averted if, as nature intended, the respirations were rigidly confined to the nasal passages, and the mouth kept comfortably though firmly shut. ... A useful precaution, therefore, in addition to exclusively breathing through the nostrils, would be to douche the nasal cavity, after a dusty run or walk, with a weak and slightly warm solution of some harmless antiseptic.

From *Nature* 14 July 1898.

50 YEARS AGO

With the progress of chemistry, and particularly of organic chemistry, it became extremely useful to denote certain definite groups of elements which pass from compound to compound apparently as units with special symbols; thus in addition to the symbols of the elements, we have 'Me' to represent the methyl group or radical, 'Ph' the phenyl group, and so on. Nevertheless, it was unusual to find such abbreviations in the text of chemical books and papers. Even more complicated substances are now being identified, particularly in the field of biochemistry, and with each new compound the urge to find and use abbreviations seems to grow. Adenosine triphosphate soon became 'ATP'; then there was 'ADP' (adenosine diphosphate); many others will come to mind. No one can blame workers in the rush of discovery jotting down abbreviations in their own notes, or even using them in writing down a reaction in symbols. But should they be used in the text of a paper?

From *Nature* 17 July 1948.

Bioremediation

Incas and alders

Peter D. Moore

The myth of the noble savage, living in close harmony with nature and peaceful coexistence with fellow creatures, has generally been abandoned by palaeo-environmentalists. Evidence from several parts of the world confirms that the environmental destruction brought about by early societies was often remarkable when considered in the light of their generally poor technological development and low population densities.

But were any of our ancestors sufficiently environmentally aware to take remedial action to protect the environment, and hence the resources upon which they depended? Pollen analytical studies of a lake in the central Peruvian Andes, carried out by Alex Chepstow-Lusty and colleagues¹, furnishes one case in which it seems that they were. From this evidence, the Inca peoples in the area may have at one time adopted a bioremediation programme to prevent soil erosion.

The arrival of prehistoric human cultures in new locations, or the development of new technologies in areas long occupied by people, has often resulted in environmental changes that we could regard as detrimental (reduced vegetation biomass, species extinctions, soil erosion). Even before the arrival of the Romans in Britain, the forests of the uplands had been stripped and replaced by moorland and blanket mires². The arrival of Polynesian peoples in New Zealand resulted in a decline of the tree genus *Dacrydium* and an expansion of bracken as forests were cleared³; the giant moas also went into decline and eventually became extinct (although the cause remains a matter of controversy). The artistically talented inhabitants of Easter Island had destroyed the jubaia palm, a vital resource for their existence, and had begun their cultural decline well before the discovery of the island by Captain Cook⁴. And so the list goes on, dispelling any notion of primitive husbandmen maintaining equilibrium with their environment.

The picture of the early inhabitants of the Peruvian Andes follows a predictably similar pattern⁵, with a deforested landscape already in existence some 4,000 years ago. According to the pollen and stratigraphical record of a small lake at 3,300 m altitude that has been investigated by Chepstow-Lusty and his colleagues¹, charcoal was abundant around 4,000 years ago, indicating the management of vegetation by burning. Grasslands were present, probably grazed by domesticated herds of camelids, and disturbed agricultural soils encouraged an abundance of

weed colonists, such as *Ambrosia*. The arable crop of the time was probably quinoa (*Chenopodium quinoa*), which cannot be identified specifically on the basis of its pollen grains because it can be confused with various weed species from the same family (Chenopodiaceae).

The abundance of pollen of Chenopodiaceae and of *Ambrosia* declines after about AD 100, which the authors explain by a general reduction in temperature that is indicated from many sources of evidence in the region, including ice cores⁶. Burning and pastoralism took over from the cultivation of quinoa, but there is also evidence for the introduction of maize as a crop during this period. This land-use system evidently proved inappropriate for the local landscape and climate, because the silting of the lake shows increasing soil erosion, and by AD 1050 the Tiwanaku civilization had collapsed.

It is at this stage that an unusual combination of events took place. The pollen record demonstrates a sudden expansion of alder trees, probably the species *Alnus acuminata*, a colonist of degraded soils. This in itself could be explained by the natural recovery of vegetation following a reduction in human activity in the area, especially as the temperature was increasing at the time. But the post-1050 period was also a time of new colonization by Inca peoples from the south, the Inca capital of Cuzco being situated only 85 km to the southeast of the study site.

Archaeological evidence suggests that these peoples constructed irrigation canals and terrace systems for maize agriculture. The combination of this increased agricultural efficiency with the expansion of *Alnus* can be explained, according to the authors, by the development of an agro-forestry system, using *Alnus* trees to stabilize slopes and reduce soil erosion. If this interpretation is correct, it is a rare example of early conservation practice, especially in the use of trees for soil stabilization.

The Spanish conquistadores arrived in the area during the 1530s and the Inca civilization went into rapid decline. The *Alnus* trees followed suit, presumably being exploited by the invading peoples as a source of firewood, but also affected by the falling temperatures of the Little Ice Age.

The seventeenth-century poet John Dryden's romantic image of the primitive lifestyle — "Free as nature first made man, ere the base laws of servitude began, when wild in woods the noble savage ran" (*The Conquest of Granada*) — may not be justified

by the general run of palaeo-environmental data. But from the Peruvian pollen record we can surmise that these mountain peoples managed their landscape in an environmentally aware and agriculturally informed manner. The studies may also present a model for the development of modern agro-forestry practices in the Andes, using native trees rather than fast-growing exotics. Our ancestors may not have been truly ecologically noble, but neither were they entirely lacking in environmental sensitivity. □

Display technology

Printing screens

Robert Wisniew

The paperless office has long been predicted as the ultimate goal of electronic information systems, but now paper is fighting back. On page 253 of this issue¹, Jacobson and colleagues at the MIT Media Laboratory describe a cheap electronic ink that can be printed on flexible substrates, including paper. The ink is a dispersion of tiny particles inside microcapsules, and it can be changed from white to black simply by applying an electric field. Its appearance approaches that of newsprint; and when combined with an addressing circuit, the electronic ink could even be used to make paper that can print on itself.

The approach of the MIT group is radically different from that of the rest of the display industry, which has concentrated on building liquid-crystal displays on glass substrates. Their new electronic ink is based on a suspension of white particles in an absorbing dye, held within 40- μ m-diameter urethane microcapsules. Because the white particles are more electrically polarizable than the dye, they can be moved around by applying an electric field. When the particles are moved to the far side of a microcapsule it appears dark; when they are on the near side it appears bright (Fig. 1).

The electronic ink is bistable: it will remain in the reflecting or absorbing state without consuming power until it is switched. And this technology is far more durable than earlier electrophoretic displays. Previously, instead of being used in an encapsulated form, electrophoretic particles have been held between two plates. Particle clustering and migration across the plates always meant that these displays became degraded after a few hundred thousand cycles, whereas the new ink has already been tested for more than ten million cycles.

To commercialize this work, a company called E Ink has been formed. Named by *Fortune Magazine*² as one of 12 "Cool Companies" for 1998, E Ink has attracted a lot of attention and \$15.8 million of investment. The company will probably concentrate on

Peter D. Moore is in the Division of Life Sciences, King's College, Campden Hill Road, London W8 7AH, UK.

e-mail: peter.moore@kcl.ac.uk

1. Chepstow-Lusty, A. J. et al. *Mountain Res. Dev.* **18**, 159–172 (1998).
2. Turner, J. in *The Environment in British Prehistory* (eds Simmons, I. G. & Tooley, M. J.) 250–281 (Duckworth, London, 1981).
3. Newnham, R. M. et al. *J. Archaeol. Sci.* **25**, 533–544 (1998).
4. Flenley, J. R. & King, S. M. *Nature* **307**, 47–50 (1984).
5. Chepstow-Lusty, A. J. et al. *Antiquity* **70**, 823–833 (1996).
6. Thompson, L. G. et al. *Science* **234**, 361–364 (1986).

producing electronic signs and other low-resolution, low-pixel-count devices (with less than 10 pixels per inch and fewer than 10,000 pixels altogether), using a separate address circuit for each pixel.

The new electronic ink is a good choice for such large-area applications because of its low cost per unit area. Whereas conventional liquid-crystal displays use two sheets of glass, which must be separated by several micrometres and made in a clean-room, the electronic ink can simply be printed on cheap materials such as paper, plastic or metal.

The key to success in other applications will be to develop higher-content displays. It is not economically feasible to directly drive each pixel in a high-content display, so instead some sort of matrix addressing must be used. Matrix addressing uses an x–y grid of wires with pixels at their crosspoints and the voltage drivers at the ends of the rows and columns (Fig. 2). This allows many pixels per drive circuit. In a simple matrix display, only the display element is located at the crosspoint; but this set-up requires a sharp voltage transition between the black and white states, and electronic ink does not have such a sharp transition. It is better suited to active-matrix techniques, where another electronic component is added at each crosspoint to

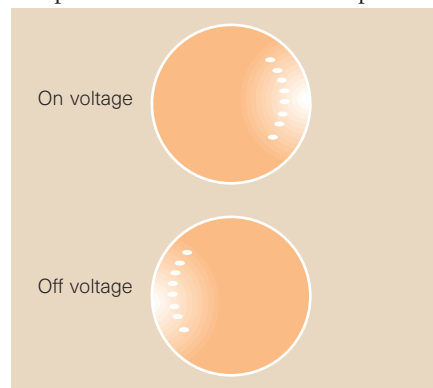


Figure 1 **Switching ink: a voltage moves encapsulated white particles to the right (turning the ink 'on', viewed from that side) or the left ('off').**

sharpen the voltage response.

Diodes are the simplest suitable components to add, and Jacobson and colleagues have already made a printed metal–insulator–metal diode display. But diode-based active matrices require extremely uniform current–voltage characteristics for all of the diodes in the array, and also small diode areas to minimize the capacitance. Transistors would require more processing steps, but would loosen the requirements of electrical uniformity and size.

What systems might be in competition with electronic ink? A lot of work has gone into liquid-crystal systems, notably in the super-twisted nematic (STN) mode. This type of display has an extremely sharp voltage transition between black and white, so simple matrix addressing can be used. It is widely used in pagers, cell phones and handheld personal digital assistants. But the reflectance of STN displays is worse than that of electronic ink, and STN displays are not bistable — so the image is lost when the power is cut. These are considerable drawbacks in many applications.

Another possibility is the cholesteric liquid-crystal mode, which can use a simple matrix-addressing scheme to address a high-content display that is bistable and has better reflectance than STN, as well as multiple colours³. But this display requires several liquid-crystal cells to be stacked, which means a more complex manufacturing method than conventional liquid-crystal displays.

In Japan, Toshiba and Sharp have exhibited prototypes and announced plans to manufacture reflective active-matrix liquid-crystal displays. The prices of these displays are expected to be comparable to existing, back-lit, colour displays that are used in notebook computers.

But none of these technologies is likely to compete for cheapness and mechanical

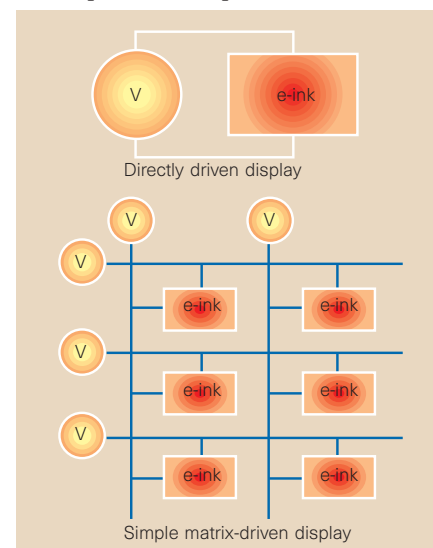


Figure 2 **To go beyond a trivial one-element display, pixels are arranged on a grid, with voltage drivers at the end of each row and column.**

flexibility with the new electrophoretic ink. In the near term, the electronic ink should allow innovations in low-resolution, low-content reflective displays such as street signs and electronic badges. In the long term, it could serve in more demanding applications—in disposable TV screens, or in active camouflage, or as electronic wallpaper, perhaps. Jacobson's long-term goal⁴ is to build complete electronic books, which can be electronically changed into any book that the reader wants. □

Robert Wisniewski is in the IBM/Thomas J. Watson Research Center, PO Box 218, Yorktown Heights, New York 10958, USA.

e-mail: wisniewski@us.ibm.com

1. Comiskey, B., Albert, J. D., Yoshizawa, H. & Jacobson, J. *Nature* **394**, 253–255 (1998).
2. Creswell, J. *Fortune Mag.* **138**, No. 1, 76–79 (1998).
3. Davis, D., Kahn, A., Huang, X. Y. & Doane, J. W. *SID 98 Digest* (ed. Morreale, J.) 901–904 (Soc. Inf. Displ., Los Angeles, CA, 1998).
4. Jacobson, J., Comiskey, B., Turner, C., Albert, J. & Tsao, P. *IBM Syst. J.* **36**, 457–463 (1997).

Autoimmunity

The pathogen connection

Christophe Benoist and Diane Mathis

Several indirect arguments support the idea that microbial agents influence the occurrence or course of certain autoimmune diseases. These include the fact that migrant populations acquire the disease prevalence of the geographical area to which they move, a prevalence correlated with latitude; that the incidence or frequency of autoimmune diseases has dramatically changed in the last two centuries; and that non-obese-diabetic (NOD) mice are protected from diabetes by bacterial infections. The nature of the agents involved and their mechanism of action remain unclear. Is there immunological cross-reactivity between anti-pathogen and anti-self responses — molecular mimicry? Or do the pathogens upset the metastable equilibrium of self-tolerance by disrupting immunoregulation — bystander activation? Recent papers^{1–3}, the latest by Horwitz *et al.*¹ in this month's *Nature Medicine*, bring interesting evidence to the debate.

Immune tolerance to self is not absolute. Given the breadth of protein epitopes in higher eukaryotes, an immune system from which all potential self-reactivity had been deleted would probably not respond to anything. Thus, clonal deletion in the thymus and peripheral lymphoid organs seems to spare a large number of T lymphocytes that are potentially reactive against self-components. These T cells essentially ignore the self-antigens because they are seen with low affinity, they occur at low abundance, or because they hide behind tissue or processing barriers.

The currently fashionable concept of molecular mimicry⁴ (Fig. 1a) proposes that pathogens express a stretch of protein that is related — in sequence or structure — to a particular self-component. This pathogen-encoded epitope can be presented by the major histocompatibility complex and activate self-reactive T cells. Activation could occur because the T cell's antigen receptor has a higher affinity for the pathogen protein than for the self-component, or because T cells are more readily primed in the inflam-

matory context of an infection. Because primed and amplified T lymphocytes have a lower threshold for activation, they can now attack self-antigens that they previously ignored.

The alternative concept of bystander activation (Fig. 1b) proposes that pathogens disturb self-tolerance without antigenic specificity coming into play. They can do this in several ways: by provoking cell death and thus the release of cellular antigens, increasing their visibility or abundance; by attracting and potentiating antigen-presenting cells; or by perturbing the cytokine balance

(either locally or over long distances) through the inflammation that is associated with infection. In a sense, bystander autoimmunity falls within the broader idea of parasite or virus-associated immunopathology — that the immune response is most deleterious to the host, not the intrinsic toxicity of the pathogen.

There is good evidence that molecular mimicry could operate. Relevant homologies between mammalian and pathogen sequences have been found (albeit in the midst of much chaff from statistically or structurally indiscriminating database searches). Experimental support has come from animals immunized with peptides containing such homologous motifs⁵, and transgenic mice in which a viral epitope is expressed on particular organs^{6,7}. But how relevant is such evidence to real-life situations in human pathology? This is a thorny issue for complex, multi-factorial diseases, because it will be particularly arduous to identify the culprit microbe if the autoimmune manifestation is a low-frequency complication of infection by common pathogens.

One case in point is Coxsackie B virus, which has been linked to autoimmune diabetes (insulin-dependent diabetes mellitus; IDDM). Sero-epidemiological evidence for such an association is sketchy⁸, but

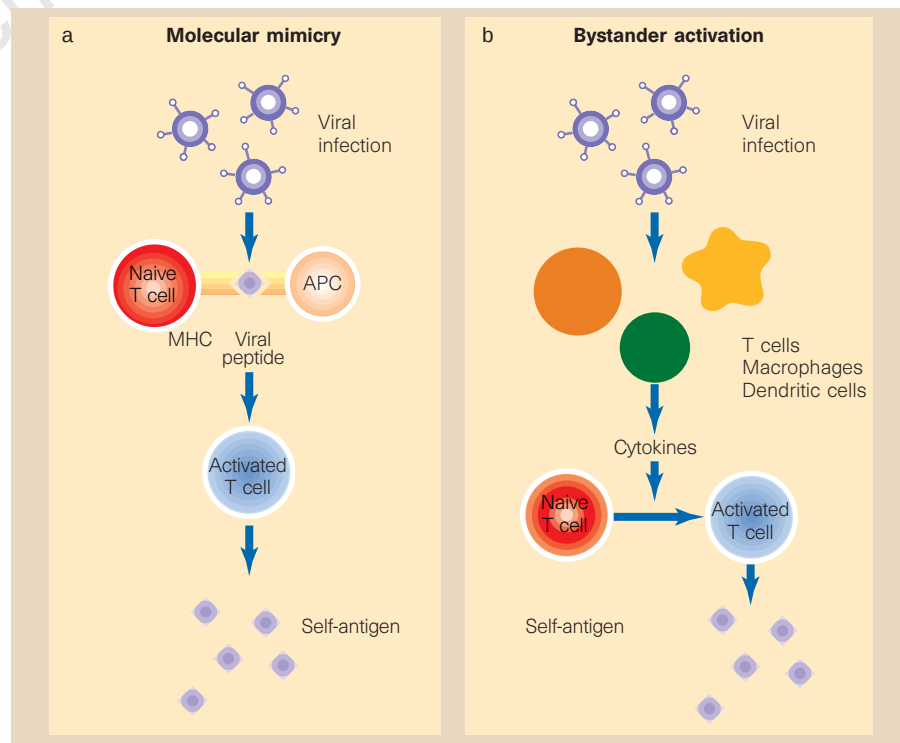


Figure 1 Two ways in which pathogens might trick the immune system into attacking its own cells. **a, Molecular mimicry.** Naive T cells recognize, and are activated by, viral antigen that is related to a self protein. When these T cells subsequently encounter that protein, they mount a cytotoxic response. **b, Bystander activation.** The virus induces an inflammatory response resulting in, among other things, the release of cytokines. Dormant T cells in close proximity are woken up, and have a lower threshold for activation, meaning that they can now attack self-antigens that they may previously have ignored. The studies of Horwitz *et al.*¹ indicate that, for Coxsackie-virus-induced diabetes at least, bystander activation may be an important mechanism.

attention has been drawn to the homology between determinants of the Coxsackie P2-C protein and glutamate decarboxylase (GAD), one of the autoantigens recognized in IDDM (for review see ref. 9). Could Coxsackie infection unleash autoreactivity to GAD and thereby provoke IDDM?

If viruses activate pathogenic autoimmunity through molecular mimicry, they should not be able to do so if the immune repertoire is blind to cross-reactive epitopes. Horwitz *et al.*¹ tested this possibility — and the potential importance of virus-induced bystander activation — by studying the BDC2.5 mouse model of diabetes. Most of the T cells in these transgenic mice are reactive against a naturally expressed pancreatic antigen that is distinct from GAD. When carried on the NOD genetic background, BDC2.5 mice show heavy infiltration of the pancreas by T cells; the local lesion is active, as shown by lymphocyte activation, division and programmed cell death, but a balance is somehow maintained such that complete destruction of insulin-producing cells is avoided for a long time¹⁰.

Horwitz and colleagues found that infection by Coxsackie B4 rapidly provoked diabetes in the transgenic mice, but not in non-transgenic littermates or in NOD animals, which show a less extensive pancreatic infiltration. This effect was at least to some degree virus specific, because it did not occur after infection by lymphocytic choriomeningitis virus. Coxsackie B4 infects pancreatic cells¹¹, so the local inflammation that it provokes probably disturbs the immunoregulatory balance of autoreactive T cells in the vicinity (increased levels of antigen and pro-inflammatory cytokines).

This interpretation is consistent with a previous analysis from the Zinkernagel group², using another transgenic system. They found that functional cytotoxic T cells could be elicited through bystander activation, but could not home to and destroy the pancreas — unlike T cells activated, in higher numbers, by recognition of cognate viral antigen. The results of Zhao *et al.*³, although interpreted in the context of molecular mimicry, also underscore the importance of local effects of pathogens. These authors found that T cells activated by a mimic from Herpes simplex virus could not provoke corneal keratitis without a local, virus-induced lesion.

Ultimately, the conclusion is that the suspected connection between Coxsackie B virus and IDDM is linked to viral infection of the pancreas and bystander activation of a pre-existing, but controlled, autoimmune state — homology to GAD would be a red herring. Although this would be overstating the case that can be made from the available data, it will be important to keep in mind these demonstrations of viral bystander effects. For example, therapeutic immunointervention focused on cross-reactive epi-

topes would be misguided if a pathogen's main contribution were bystander activation of dormant autoreactive cells. □

Christophe Benoist and Diane Mathis are at the Institut de Génétique et de Biologie Moléculaire et Cellulaire (CNRS/INSERM/ULP) BP 163, 67404 Illkirch, CU de Strasbourg, France.

e-mail: cbdm@igbmc.u-strasbg.fr

1. Horwitz, M. S. *et al.* *Nature Med.* **4**, 781–785 (1998).
2. Ehl, S. *et al.* *J. Exp. Med.* **185**, 1241–1251 (1997).

3. Zhao, Z.-S. *et al.* *Science* **279**, 1344–1347 (1998).
4. Oldstone, M. B. *Cell* **50**, 819–820 (1987).
5. Fujinami, R. S. & Oldstone, M. B. *Science* **230**, 1043–1045 (1985).
6. Ohashi, P. *et al.* *Cell* **65**, 305–317 (1991).
7. Oldstone, M. B. *et al.* *Cell* **65**, 319–331 (1991).
8. Graves, P. M. *et al.* *Diabetes* **46**, 161–168 (1997).
9. Ellis, T. M. & Atkinson, M. A. *Nature Med.* **2**, 148–153 (1996).
10. André, I. *et al.* *Proc. Natl Acad. Sci. USA* **93**, 2260–2263 (1996).
11. Notkins, A. L. *Arch. Virol.* **54**, 1–14 (1977).

Oceanography

Eddies make ocean deserts bloom

Richard G. Williams and Michael J. Follows

Phytoplankton require nutrients for growth and reproduction, and they need to have them in the euphotic zone, the sunlit 'skin' of the ocean where photosynthesis can take place. Traditionally, nutrient supply to the euphotic zone has been explained in terms of large-scale circulation over entire ocean basins. However, three papers^{1–3}, two of them^{2,3} in this issue, invoke a complementary explanation — that energetic eddies, analogous to atmospheric weather systems, are a source of nutrients on a much smaller horizontal scale of several tens to hundreds of kilometres.

The central question to be answered is as follows. How is phytoplankton primary production sustained at observed levels in nutrient-poor parts of the ocean? Satellite observations⁴ show that primary production is greatest at high latitudes and near coastal boundaries (Fig. 1). These areas broadly correspond with regions where winds induce upwelling of deep, nutrient-rich waters. Primary productivity is still observed over the comparatively nutrient-depleted subtropical gyre — the 'ocean desert' at mid-latitudes, which is characterized by downwelling (shown by the negative contours in Fig. 1). That in itself is no surprise because much of primary production uses recycled nutrients. But a fraction of primary production — new production — is sustained by newly supplied nutrients balancing a loss from sinking organic particles.

So where do these nutrients come from? Traditional explanations are unconvincing. Vertical diffusion is not the answer — estimates of vertical mixing rates⁵ in the upper thermocline, where there are strong vertical gradients, are an order of magnitude too small to balance the nutrient budget. Atmospheric input of nitrogen is significant near the coast, but relatively small over the centre of the subtropical gyre. Lateral input of nitrate (one of the principal nutrients used by phytoplankton) from neighbouring upwelling regions occurs along the flanks of the gyre, but this contribution peters out towards the centre⁶.

Last year, McGillicuddy and Robinson¹ proposed that in the Sargasso Sea, a sub-

region of the North Atlantic subtropical gyre, the time-varying eddy field might supply the required nutrients. They argued that geostrophic eddies cause nutrients to be 'upwelled' and 'downwelled' through the base of the euphotic zone. Upwelled nutrients are partly consumed by phytoplankton in the euphotic zone and converted into organic matter, whereas there is no equivalent conversion for downwelled nutrients. So eddies result in a rectified transport of nutrients into the euphotic zone because of the asymmetry in the ecosystem response; this process is analogous to how a combination of atmospheric eddy transport and photochemistry results in a maximum in total ozone at mid-to high latitudes in the atmosphere.

In the first of the papers in this issue (page 263), McGillicuddy *et al.*² extend this work — they discuss how new production levels of up

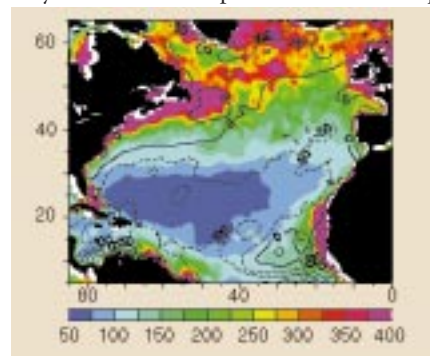


Figure 1 Annual primary productivity (coloured values in $\text{g C m}^{-2} \text{yr}^{-1}$), and upwards vertical velocity of water (contoured in m yr^{-1}) in the North Atlantic. Productivity reaches maximum values of $400 \text{ g C m}^{-2} \text{yr}^{-1}$ where there are high levels of nutrient input from upwelling; and it has minimum values of $50 \text{ g C m}^{-2} \text{yr}^{-1}$ within the subtropical gyre where there is downwelling (negative contours) and comparative nutrient depletion. Previous estimates of nutrient supply seem inadequate to account for even these low values, hence the proposal^{1–3} that eddy circulation may be responsible for supplying them. (Figure derived from satellite estimates of surface chlorophyll from ref. 4, and calculations of vertical velocity at the base of the surface wind-forced boundary (Ekman) layer⁶.)

to $0.50 \pm 0.14 \text{ mol N m}^{-2} \text{ yr}^{-1}$ in the Sargasso Sea, off Bermuda, might be partly maintained by an eddy nitrate supply of $0.35 \pm 0.1 \text{ mol N m}^{-2} \text{ yr}^{-1}$ (ref. 1) or $0.19 \pm 0.1 \text{ mol N m}^{-2} \text{ yr}^{-1}$ (ref. 2) (the estimates come from two different methods). In the second (page 266), Oschlies and Garçon³ describe the results of their investigations of this process over the whole North Atlantic basin. They use an eddy-resolving, general-circulation model, which includes a simplified ecosystem model and assimilation of altimetric data to make the eddy field more realistic. Their estimates of nutrient supply are in accord with observed levels of new production in the Sargasso Sea, but fall significantly beneath those required to sustain primary production over other parts of the subtropical gyre.

This underestimate might be partly because the resolution of the model is still too coarse to include vigorous upwelling at fronts on horizontal scales of a few tens of kilometres. But new production in the subtropical gyre might also be sustained by other nitrogen sources, such as fixation of dissolved nitrogen gas, or transport and remineralization of organic nitrogen. This view is supported by the poleward nitrate flux identified across 36° N in the Atlantic, which requires another source to close the nitrogen budget⁷.

Nitrogen fixation may enhance new production, with the resulting organic matter increasing the nitrate:phosphate ratio upon remineralization at depth. Such a process appears to be occurring in upper thermocline waters of the North Atlantic^{8,9}. The implied nitrogen fixation rate is estimated to be $0.07 \text{ mol N m}^{-2} \text{ yr}^{-1}$ averaged over the Sargasso Sea⁹ (but note that this figure is smaller than that for the eddy supply of nitrate^{1,2}).

In our view, transport and remineralization of dissolved organic nitrogen (DON) might be particularly important in the subtropical gyre, because DON becomes the largest pool of reactive nitrogen in surface, nitrate-depleted waters⁸. New production might be enhanced here through the lateral supply of DON from productive, upwelling zones⁶. This nutrient supply might penetrate further into the subtropical gyre than does nitrate, because the semi-labile pool of DON has a longer lifetime in the euphotic zone. However, there are insufficient observations of DON in the Atlantic to test this hypothesis.

In a wider context, time-series observations in the Sargasso Sea¹⁰ reveal an unexplained loss of dissolved inorganic carbon (DIC) in the upper 150 m during summer. A summertime supply of nutrients could enhance new production and provide an added drawdown of DIC, which is used during photosynthesis. Although the eddy-pumping mechanism of McGillicuddy *et al.*² can explain the levels of new production off Bermuda, it does not lead to a corresponding decrease in DIC because enough extra carbon is upwelled from depth to offset the increased

consumption during photosynthesis.

Taken together, these studies bring us closer to understanding how 'ocean deserts' bloom. The remaining challenge is to identify the relative importance of the various processes involved. □

Richard G. Williams is in the Oceanography Laboratories, University of Liverpool, Liverpool L69 3BX, UK.

e-mail: ric@liv.ac.uk

Michael J. Follows is in the Program in Atmospheres, Oceans and Climate, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

e-mail: mick@plume.mit.edu

Ecology

Rarity as double jeopardy

Kevin J. Gaston

In the introduction to his book *Geographical Ecology*¹, published over 25 years ago, Robert H. MacArthur opened with the words, "To do science is to search for repeated patterns, not simply to accumulate facts". Since then, the roll-call of such ecological patterns has steadily grown. Indeed, one of the most pervasive at broad geographical scales — and the subject of a paper by Johnson² on page 272 of this issue — makes no appearance in MacArthur's book.

The pattern in question is this. For a given taxonomic assemblage at a geographical scale — say, the terrestrial mammals of Britain or the breeding birds of North America — the more abundant species tend also to be the more widely distributed^{3–5}. This makes sense as more individuals require, on average, a larger area. However, the numbers of individuals actually increase at a faster rate than does the area over which they are distributed, such that more widely distributed species not only tend to have larger numbers of individuals but also usually occur at higher local densities. In other words, there is a positive relationship between the average local densities and the geographical distributions of species in an assemblage.

Low local abundances and small geographical ranges typically increase the risk of extinction that species face. Local rarity increases the likelihood that demographic and environmental stochasticity will wipe out populations, and a restricted distribution means that all or most individuals will probably experience adverse conditions simultaneously. The compilation of lists detailing those species that are most threatened with regional or global extinction (the so-called Red Lists, such as that recently published for plants globally⁶) relies heavily on these observations. The positive interspecific relationship between local abundance and geographical distribution means that, with regard to risk of extinction, species will tend

1. McGillicuddy, D. J. Jr & Robinson, A. R. *Deep-Sea Res.* **1** **44**, 1427–1449 (1997).
2. McGillicuddy, D. J. Jr *et al.* *Nature* **394**, 263–266 (1998).
3. Oschlies, A. & Garçon, V. *Nature* **394**, 266–269 (1998).
4. Sathyendranathan, S., Longhurst, R. S. A., Caverhill, C. M. & Platt, T. *Deep-Sea Res.* **1** **42**, 1773–1802 (1995).
5. Ledwell, J. R., Watson, A. J. & Law, C. S. *Nature* **364**, 701–703 (1993).
6. Williams, R. G. & Follows, M. J. *Deep-Sea Res.* **1** **45**, 461–489 (1998).
7. Rintoul, S. R. & Wunsch, C. *Deep-Sea Res.* **1** **38**, S355–S377 (1991).
8. Michaels, A. F. *et al.* *Biogeochemistry* **35**, 181–226 (1996).
9. Gruber, N. & Sarmiento, J. L. *Glob. Biogeochem. Cycles* **11**, 235–266 (1997).
10. Michaels, A. F., Bates, N. R., Buessler, K. O., Carlson, C. A. & Knap, A. H. *Nature* **372**, 537–540 (1994).

to face a 'double jeopardy'⁷ — if they occur at low abundance they are likely also to be narrowly distributed. This obviously has profound consequences for conservation.

Johnson² now provides an elegant demonstration of this double jeopardy. He found that for marsupials on the Australian mainland there is a positive, albeit rather weak, relationship between the local density of species and the extent of their geographical occurrence. At first sight, Johnson's finding seems to be rather unsurprising. But he then separated the species into two groups (Fig. 1, overleaf) — those of recent evolutionary origin (less than 4 million years old) and those that are more ancient (assuming that species with no recent surviving relatives are genuinely old, and do not simply seem to be so because their relatives have become extinct quickly⁸).

Johnson found that recently evolved species show a positive relationship between local abundance and geographical spread that is much stronger than that documented across the entire assemblage. The more ancient species, by contrast, show a negative relationship — that is, locally abundant species tend to be narrowly distributed and locally rare species tend to be more widespread. Johnson argues that such a result might be expected if, over evolutionary time, species that are both locally rare and narrowly distributed have a disproportionate likelihood of becoming extinct. In other words, those that face double jeopardy suffer the consequences, and the relatives of the ancient species that had low local density and small range have been lost.

The finding that by considering differing subsets of species we can influence our interpretation of the relationships observed, brings to mind an influential study of another macroecological pattern by Nee *et al.*⁹. For breeding birds in Britain they found that, although abundance declined with body size



Figure 1 'Recent' and 'ancient' species of Australian marsupial. Left, *Macropus giganticus*, a recent species; right, *Phascolarctos cinereus*, an ancient species. Johnson² has found that species of relatively recent evolutionary origin (less than 4 million years old) show a positive relationship between local abundance and geographical spread, and more ancient species a negative one. Presumably, this is because the relatives of ancient species who showed low abundances and small ranges have already become extinct.

across all species, within individual tribes (such as woodpeckers) the less recent the common ancestry they shared with other British birds the more likely the relationship was to be a positive one. The six tribes that were most distantly related to other tribes all showed positive abundance–body-size relationships.

Considerable insights can be gained into the effects of processes that operate over evolutionary timescales on contemporary ecological patterns, simply by doing separate analyses for species of different ages. Johnson's study also illustrates the power of the 'comparative method' in revealing patterns in ecological data^{10,11}. The analyses are controlled for the non-independence of data points that results from the evolutionary relatedness of species. This at a time when the usefulness of the comparative method in ecology, rather than evolutionary biology itself, is being much debated.

The relative importance of local abundance and geographical distribution in determining risks of global extinction, and the independence of these effects, remains unknown. Johnson's analyses indicate a surprisingly simple conclusion in this regard. The slope of the lower boundary to the negative relationship between local abundance and range size for the ancient species is approximately -1 (see Fig. 2b on page 273). This implies that greater local abundance or greater range size compensate almost exactly for a reduction in the other variable. If it proves to be generally true, this will be a very significant result and might usefully simplify considerations of extinction risk.

The analyses, as reported, do not reveal whether surviving ancient species of Australian marsupials face any greater risk of extinction than the more recently evolved

species. However, a higher proportion of the ancient species appear to be listed as at a significant risk of extinction in the near future, according to the current IUCN Red List of threatened animals¹². This is despite the fact that the mean densities and range sizes of the ancient species are not obviously lower than those of the recent species, suggesting that age or some correlate of age may be causal (or that the likelihood of listing is biased by age or taxonomic distinctiveness). Other studies have also found that older species — or at least species in older groups — tend to be at greater risk^{13,14}.

Having established the link between science and the search for repeated patterns, MacArthur¹ continued, "many take refuge in nature's complexity as a justification to oppose any search for patterns". Along with other recent papers in the field of macroecology, that by Johnson makes the refuge seem an increasingly dull place to be. □

Kevin J. Gaston is in the Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK.

e-mail: k.j.gaston@sheffield.ac.uk

1. MacArthur, R. H. *Geographical Ecology: Patterns in the Distribution of Species* (Harper & Row, New York, 1972).
2. Johnson, C. N. *Nature* **394**, 272–274 (1998).
3. Hanski, I. *Oikos* **38**, 210–221 (1982).
4. Brown, J. H. *Am. Nat.* **124**, 255–279 (1984).
5. Gaston, K. J. *Oikos* **75**, 211–220 (1996).
6. Walter, K. S. & Gillett, H. J. (eds) 1997 *IUCN Red List of Threatened Plants* (IUCN, Gland, Switzerland, 1998).
7. Lawton, J. H. *Trends Ecol. Evol.* **8**, 409–413 (1993).
8. Gaston, K. J. & Blackburn, T. M. *Proc. R. Soc. Lond. B* **263**, 63–68 (1996).
9. Nee, S., Read, A. F., Greenwood, J. J. D. & Harvey, P. H. *Nature* **351**, 312–313 (1991).
10. Rees, M. *J. Ecol.* **83**, 891–892 (1995).
11. Harvey, P. H. *J. Anim. Ecol.* **65**, 255–263 (1996).
12. IUCN 1996 *IUCN Red List of Threatened Animals* (IUCN, Gland, Switzerland, 1996).
13. Gaston, K. J. & Blackburn, T. M. *Evol. Ecol.* **11**, 557–565 (1997).
14. McDowall, R. M. *Tuatara* **17**, 1–12 (1969).

Daedalus

His master's fist

Human style is a very subtle concept. Each of us has a characteristic 'house style' which shows up in everything we do. Even a Morse code operator, who can merely switch a circuit between two states, develops a personal 'fist' that can be recognized by other operators. Machines which could recognize their operator's 'fist' would be almost impossible to hijack or misuse. So Daedalus is inventing them.

The personal computer is the obvious first choice for the treatment. DREADCO's prototype has a special neural-net learning program. It accumulates its owner's keyboard 'fist', his mouse style and his characteristic pattern of usage. Soon it will recognize him very clearly. A stranger seeking to explore its files will be refused access, even if he has stolen the correct password. He will merely download a stream of abuse, together with threats that the machine will know him next time — which of course it will. Meanwhile, it will still respond to its owner, even from keyboards miles away on the network.

A computer thus truly 'personalized' might even get to recognize when its owner was dead tired, or drunk, or in one of his moods again. It might refuse him access to particularly delicate files, or even tell him to go and sleep it off, and shut down.

A machine used by several people, as in a busy office, will take slightly longer to recognize them all. New employees will have to be 'inducted' by someone it has already learnt to trust. The slight initial complications will be well repaid by the subsequent improvement in security.

Cars, too, could benefit from being personalized. A car whose neural net accepted inputs from the steering wheel, accelerator, gear shift and all other movables, would soon learn the style of its driver. It would know how widely he opens the door to get in, at what speeds he changes gear, with what angular increments he corrects steering deviations, and so on. A strange hand on the wheel would at once arouse its suspicions. Further alien actions would rapidly confirm them. It would stop dead and, through the stereo speakers, shout that it was being stolen.

Now that nearly all human artefacts have a microprocessor inside them, machine recognition could spread rapidly. The bank auto-teller, the TV, the fridge, even the house front door, all could learn to greet their approved users, while snarling at attempted interlopers. The 'Haves' will be even safer from the 'Have-nots'.

David Jones

Feynman's figurations

If a picture is worth a thousand words, a diagram can be worth many lines of complicated algebraic formulas. Feynman diagrams are also powerful tools of explanation and prediction.

Martin Kemp

How are we to understand Richard Feynman's statement, repeated in various forms, that "nobody understands quantum mechanics"? As the great authority on quantum electrodynamics and an eloquent communicator of difficult physics to nonspecialist audiences, his aphorism cannot be taken at face value.

He could not have meant that it was impossible to comprehend the basic principles, nor was he intending to dismiss the analytical and predictive efficacy of the mathematics of quantum mechanics. Rather, his assertion has to be interpreted in the light of the gravitational pull of physical modelling and graphic picturing in his approach to the physics.

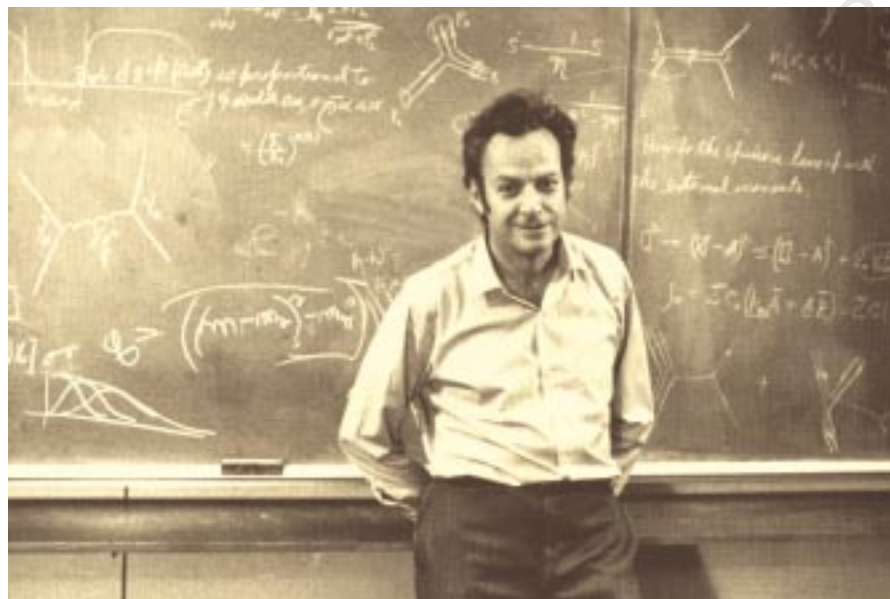
He was continually concerned about the way that physical behaviour could be expressed in algebraic conventions without concrete visualization: "Strange! I don't understand how it is that we can write mathematical expressions and calculate what the thing is going to do without being able to picture it."

As Freeman Dyson recalled, "he was a natural physicist, and he thought in terms of concrete objects. The mathematics was an encumbrance, something you had to stick on afterwards, more or less ... as a necessary evil."

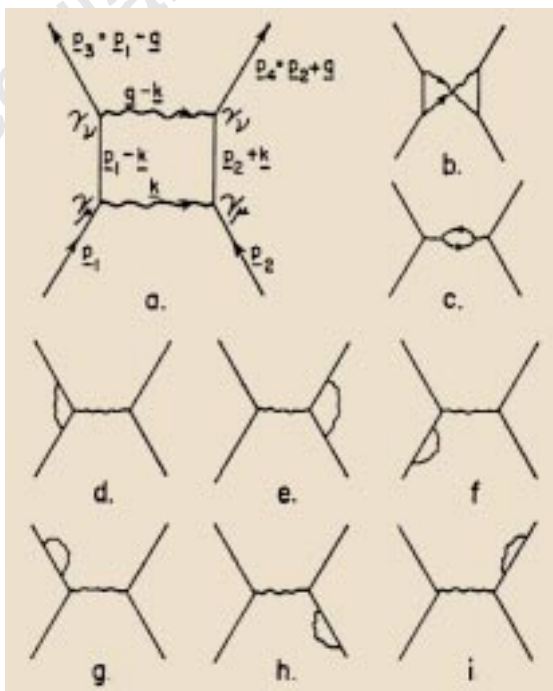
The now ubiquitous 'Feynman diagrams' achieved the graphic encoding of phenomena whose very nature means that they cannot be subject to direct picturing. In the 1949 paper that inaugurated his method, he provided the classic demonstration of the "fundamental interaction" in which electrons exchange a proton. The electrons are figured as solid lines, while the process of exchange of a 'virtual quantum' is represented by a wavy line, which graphically signals that the photon can be emitted from one electron and absorbed by the other, or if it travels backwards in time, in reverse. Either way, the result is the deflection of both electrons, denoted by their angular deviation.

Feynman diagrams look superficially like the simple graphics that physicists have used for centuries. But they are devices of exceptional power. Within their space-time coordinates, Feynman was able to sidestep the long-winded algebraic formulas that treated electrons and positrons separately. All the equations came together in one picture in a way that preceded and even directed calculation.

The diagrams rapidly and economically



Feynman and his diagrams (from *No Ordinary Genius: The Illustrated Richard Feynman*, edited by Christopher Sykes; Weidenfeld and Nicolson, 1994).



Feynman diagrams of "the interaction between two electrons" (from "Space-Time Approach to Quantum Electrodynamics", *Physical Review* 76, 769–789; 1949).

explained and predicted in ways that were both intuitive and analytical. Feynman himself described the intuitive element in a characteristically graphic way: "It's like asking a centipede which leg comes after which — it happens quickly, and I'm not exactly sure what flashes and things go on in the head. I do know it's a crazy mixture of partially solved equations and some kind of visual picture of what the equation is saying is happening, but not as well separated as the words I'm using."

The diagrams mirror his conviction that

"there is ... a rhythm and a pattern between the phenomena of nature which is not apparent to the eye, but only to the eye of analysis".

Such a potent grammar of diagrams and matching equations provides a marvellous tool. Yet is it reasonable to wonder whether it also constrains what minds lesser than Feynman's permit themselves to envisage? □

Martin Kemp is in the Department of the History of Art, University of Oxford, 35 Beaumont Street, Oxford OX1 2PG, UK.

Selection on swallow tail streamers

Sexual selection results when one sex, usually females, discriminates among members of the opposite sex during mate choice. It is interesting because it is a mechanism by which those traits used in mate choice can become exaggerated in a way that is apparently detrimental, except in terms of mating, to the bearer of the ornaments¹. One of the classic examples of an exaggerated trait produced by sexual selection is the tail streamers of swallows (*Hirundo rustica*)². Here, however, by using tail-manipulation experiments, I show that the tail streamers of swallows are mainly the product of natural selection, with sexual selection only causing the streamer to extend beyond the aerodynamic optimum by a small amount.

Sexually selected traits give individuals an advantage over members of the same sex in access to mates of the opposite sex^{1,3}. In practice, however, it is not straightforward to distinguish between the products of sexual and natural selection because many traits are, to some extent, influenced by both. Swallows have been the focus of a long-term study on sexual selection³. Møller and colleagues have shown that females prefer males with elongated tails as mates and discriminate against males with shortened tails^{2,4}, suggesting that the long

tail is a product of sexual selection. But a recent aerodynamic analysis indicated that the swallow's tail streamers may function as a flight control device⁵, which implies that they could result from natural selection.

Naturally selected structures are predicted to be at an optimum level that maximizes the net benefit of the trait. If the streamer is at a naturally selected optimum, then deviations away from this optimum should increase costs; in other words, both elongating and shortening the streamer should increase flight costs.

On the other hand, a sexually selected ornament should be costly at stable equilibrium¹. Therefore, if the streamer is the product of sexual selection it should be a costly trait (as it has been extended beyond the naturally selected optimum for flight) and costs should rise as streamer length is increased and drop as streamer length is reduced.

A third possibility is that the streamer is a naturally selected structure which sexual selection has modified away from the aerodynamic optimum. If this is the case then costs should rise as streamer length is increased, but if streamer length is reduced costs should drop until the naturally selected optimum is reached and then rise as the tail is shortened away from the optimum⁶.

The crucial manipulation that distinguishes between the 'natural selection alone' and 'sexual selection alone' hypotheses is the reduction manipulation. I predicted that there would be increased costs after streamer reduction if the natural selection hypothesis were true, and decreased costs if the sexual selection hypothesis were correct.

However, it is not obvious what direction of change would represent an increase in costs. For example, is it better to fly quickly or slowly? Is it better to use large or small turning circles? The criterion that I used was to ask whether streamer elongation and reduction changed a flight parameter in the same direction relative to controls. Therefore, if the reduction manipulation resulted in significant changes in the flight parameter in relation to controls and in the same direction as the changes produced by the elongation manipulation, this would falsify the sexual selection alone hypothesis and support the natural selection alone hypothesis.

Male swallows were caught in 1995/96 and assigned randomly to four manipulation groups (tail elongated or reduced by 20 mm, unmanipulated, or manipulated but with no change in length). The 20-mm manipulation was done in the basal portion of the tail in 1995, and in the (more sexually dimorphic⁶) thin streamer in 1996. This means that the year and manipulation type (that is, basal versus streamer) effects are potentially confounded. Tail manipulations were done by J. Aparicio and A. Hall. Length of the outer tail feathers did not vary between the two years of the study ($F_{1,24}=0.83$, not significant) or among manipulation groups ($F_{3,23}=2.38$, not significant).

I filmed the birds using stereo-video when they were feeding their nestlings. I digitized the stereo-video to obtain the coordinates of the filmed bird in each frame of both videos. The filming and digitizing were conducted blind to the manipulation of the bird. From the reconstructed three-dimensional flight path of the bird, flight parameters could be calculated⁷. This technique was initially developed in collaboration with A. Thomas. The flight parameters used were velocity (mean, maximum and minimum), acceleration (mean, maximum and minimum), agility (rate of change of heading: mean and maximum) and turn radius (mean and minimum). All ten flight-performance parameters except maximum agility showed significant repeatability⁸.

Significant correlations among the ten flight variables indicated that they did not vary independently. Multiple regression identified the combination of mean

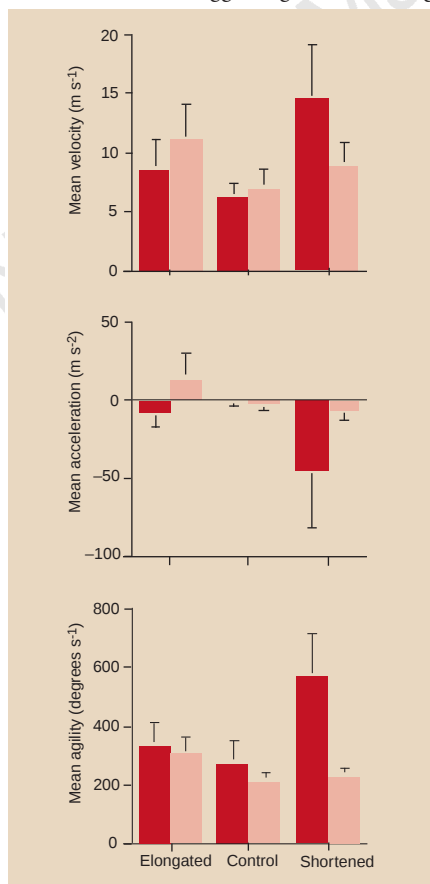


Figure 1 Relationship between manipulation group and top, mean flight velocity, centre, mean acceleration, and bottom, mean agility. Data are mean $\pm$ standard errors for the birds in each group. Red bars represent streamer manipulations done in 1996; pink bars represent basal manipulations done in 1995. Sample sizes are: elongated: streamer 5, basal 9; reduced: streamer 4, basal 9; controls: streamer 8, basal 8. Analyses of variance using the general linear model revealed significant interactions between manipulation group and manipulation type/year for all three parameters (mean flight velocity, $F_{2,42}=5.76$, $P<0.05$; mean acceleration, $F_{2,42}=7.36$, $P<0.05$; mean agility, $F_{2,42}=5.41$, $P<0.05$). The significant interaction indicates that there were significant effects of both manipulation direction and manipulation type/year and that the effect of manipulation direction is different for the two manipulation types or years. There were also significant effects of the farm on which the birds were filmed in each of the three analyses; this probably relates to the different physical arrangements of the farms. None of the morphological variables (original streamer length, wing length, or body mass) measured explained the significant variation in flight performance.

velocity, mean acceleration and mean agility as the smallest set of parameters that explained significant variation in all the others. I therefore used these as independent measures of flight performance. These three variables were not significantly correlated with each other.

For all three of these independent flight parameters, the manipulation group and the interaction between manipulation group and manipulation type/year significantly influenced flight performance. The reduction and elongation manipulations influenced performance in the same direction for all three parameters in the streamer-manipulation experiment (1996) and for two of the three parameters in the basal manipulation experiment (1995) (Fig. 1). This produces the 'U-shaped' pattern (with fitness costs on the y axis plotted against manipulated-streamer length on the x axis) predicted by the natural selection alone hypothesis and is different to the positive slope predicted by the sexual selection alone hypothesis. As it is the streamer rather than the basal part of the tail that shows significant sexual dimorphism², the results from the streamer manipulation are probably more biologically meaningful⁶.

The results of this study oppose the predictions made from the sexual selection alone hypothesis. The tail streamers of male swallows could have evolved through natural selection alone but cannot have evolved entirely through sexual selection. However, it is possible that sexual selection has been involved in elongating the tail by less than the 20-mm manipulation used. Therefore, sexual selection could be responsible for the sexual dimorphism seen in this species (streamer length differs by about 15 mm between the sexes²).

The results of Møller and colleagues indicate that sexual selection may be operating at some level on male tail-feather length. My results indicate that sexual selection is probably responsible only for increasing male tail length beyond a naturally selected aerodynamic optimum; sexual selection can be responsible for a maximum of 27% of the streamer length.

Matthew R. Evans

Department of Biological and Molecular Sciences,
University of Stirling, Stirling FK9 4LA, UK
e-mail: MRE2@stir.ac.uk

1. Andersson, M. *Sexual Selection* (Princeton Univ. Press, Princeton, 1994).
2. Møller, A. P. *Sexual Selection and the Barn Swallow* (Oxford Univ. Press, Oxford, 1994).
3. Darwin, C. *The Descent of Man, and Selection in Relation to Sex* (Murray, London, 1871).
4. Smith, H. G. & Montgomerie, R. *Behav. Ecol. Sociobiol.* **28**, 195–201 (1991).
5. Norberg, R. A. *Proc. R. Soc. Lond. B* **257**, 227–233 (1994).
6. Evans, M. R. & Thomas, A. L. R. *Proc. R. Soc. Lond. B* **264**, 211–217 (1997).
7. Rayner, J. M. V. & Aldridge, H. D. J. *N. J. Exp. Biol.* **118**, 247–265 (1985).
8. Lessells, C. M. & Boag, P. T. *Auk* **104**, 116–121 (1987).

The oldest coelurosaurian

We report here the discovery of a therizinosaur from the Early Jurassic Lower Lufeng Formation (Sinemurian stage) of Yunnan, China. This discovery extends the age range of these unusual animals, previously known only from the Cretaceous period (Albian–Maastrichtian stage¹), back by another 94 million years. This is the oldest definitive record of a coelurosaurian theropod, which therefore minimizes the divergence time for members of the group. Most important, it contradicts the

theory that the non-avian Coelurosauria occur too late in the fossil record to be related to birds.

This specimen (V11579, held at the Institute of Vertebrate Paleontology and Paleoanthropology, Beijing), consisting of most of the left dentary and part of the splenial (Fig. 1), was collected from the bottom of dull, purplish beds of the Lower Lufeng Formation of Eshan County, Yunnan. The Lower Lufeng Formation was initially thought to have been formed in the Late Triassic² but it is now widely accepted to be of Early Jurassic origin on the basis of the invertebrate (such as ostracod, pelecypod and gastropod) and vertebrate (for example, prosauropod and the tritylodontid *Bienotherium*) fossils it contains³.

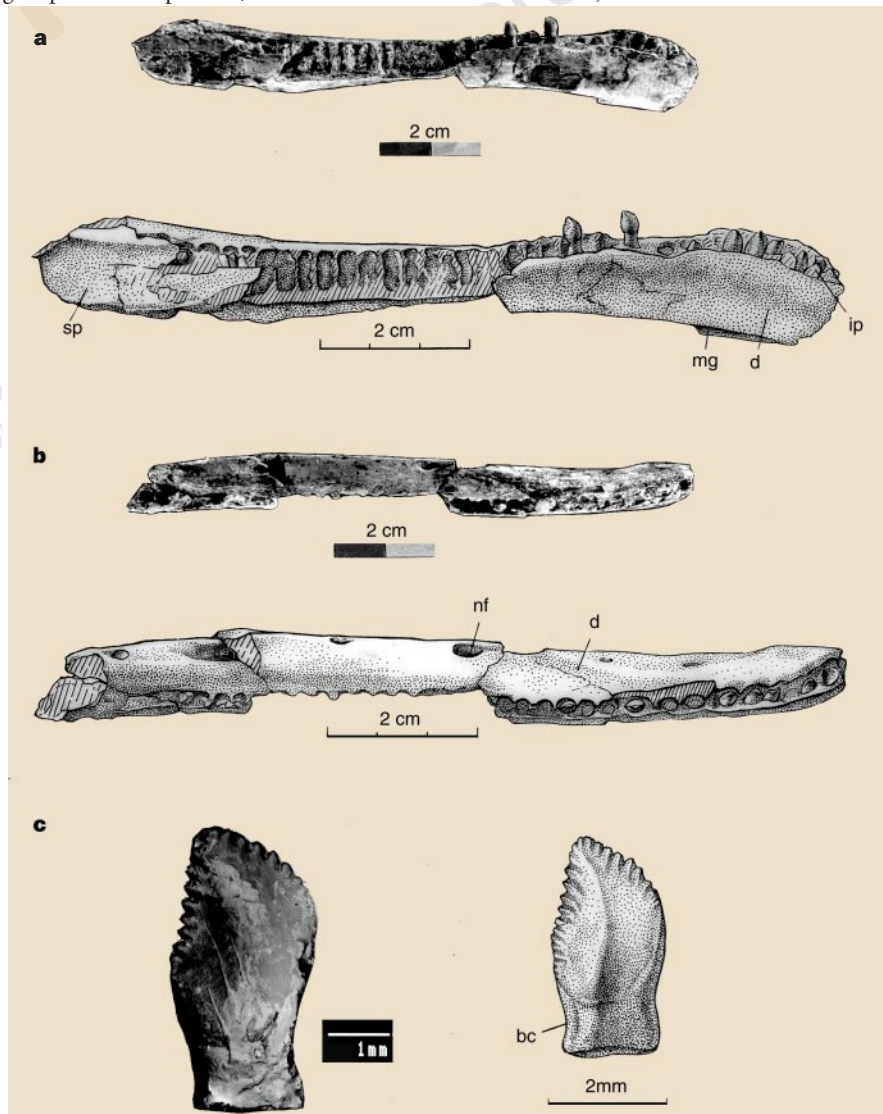


Figure 1 The left mandible of the specimen (V11579). **a**, Medial view. The downturned dentary and numerous small teeth are shared with therizinosaurids and a few other coelurosaurian theropods (troodontids and ornithomimosaurs), whereas the posterior decrease in size of the teeth and the morphology of the teeth are unique to therizinosaurids. **b**, The dorsal view shows a lateral shelf on the dentary. This is also present in ornithischians but is never so broad and flat as in this specimen and therizinosaurids. **c**, Tooth crown in lateral view. The tooth morphology, although superficially similar to that of some ornithischians and basal sauropodomorphs, is also unique to therizinosaurids in its combination of a basal constriction, circular roots, and lanceolate crown. bc, basal constriction; d, dentary; ip, interdental plate; mg, meckelian groove; nf, nutrient foramina; sp, splenial.

The therizinosauroid affinities of the specimen are indicated by the presence of a broad and flat shelf lateral to the tooth row and the few, large nutrient foramina below it; the downturned anterior end of the dentary; the decrease in the size of teeth posteriorly; the large number and small size of the teeth; and the morphology of the teeth (Fig. 1).

Therizinosauroidea ('segnosaurs') are a poorly known group of Asian dinosaurs with an unusual combination of features that, until recently, obscured their relationships. Recent hypotheses link segnosaurs with therizinosaurid theropods¹ and the group as a whole is placed in the Coelurosauria^{1,4}, which includes birds and all other theropods more closely related to birds than to Carnosauria⁵.

Previously, the oldest definitive records of Coelurosauria were *Ornitholestes* and *Coelurus* from the Late Jurassic Morrison Formation and *Compsognathus* from the Late Jurassic of Germany and France⁶. The oldest of these specimens are from beds correlated with the Kimmeridgian marine stage. Isolated teeth tentatively identified as dromaeosaurid have been found in Middle Jurassic deposits⁷. The Late Triassic *Protoavis texensis* may also belong to this group⁸, but questions have been raised about the association and interpretation of the material⁹. The new specimen is therefore the oldest fossil definitively referable to Coelurosauria, extending the record of this group back by roughly 30 million years.

The fossil record of Coelurosauria has been interpreted as inconsistent with a close relationship between birds and other members of this group¹⁰, in that most of the diversity of the non-avian Coelurosauria occurs much later in the fossil record than *Archaeopteryx* in the Late Jurassic. The presence of a non-avian coelurosaurian in the Early Jurassic indicates that by the Late Jurassic the major clades of this group must have already diverged, well before *Archaeopteryx* appears in the fossil record.

Xijin Zhao, Xing Xu

*Institute of Vertebrate Paleontology and
Paleoanthropology, Chinese Academy of Sciences,
142 Xizhimenwai Street, Beijing 100044,
People's Republic of China
e-mail: xxu@sun.midwest.com.cn*

1. Russell, D. A. & Dong, Z. *Can. J. Earth Sci.* **30**, 2107–2127 (1993).
2. Bien, M. N. *Bull. Geol. Soc. China* **20**, 225–234 (1940).
3. Luo, Z. & Wu, X.-C. in *In the Shadow of the Dinosaurs* (eds Fraser, N. C. & Sues, H.-D.) 251–270 (Cambridge Univ. Press, New York, 1994).
4. Clark, J. M., Perle, A. & Norrell, M. A. *Am. Mus. Novit.* **3115**, 1–39 (1994).
5. Currie, P. J. in *Encyclopedia of Dinosaurs* (eds Currie, P. J. & Padian, K.) 241 (Academic, San Diego, 1997).
6. Norman, D. B. in *The Dinosauria* (eds Weishampel, D. B., Dodson, P. & Osmolska, H.) 280–305 (Univ. California Press, Berkeley, 1990).
7. Evans, S. E. & Milner, A. R. in *In the Shadow of the Dinosaurs* (eds Fraser, N. C. & Sues, H.-D.) 303–321 (Cambridge Univ. Press, New York, 1994).

8. Chatterjee, S. *Phil. Trans. R. Soc. Lond. B* **332**, 277–346 (1991).
9. Ostrom, J. H. *Archaeopteryx* **14**, 39–42 (1996).
10. Feduccia, A. *The Origin and Evolution of Birds* (Yale Univ. Press, New Haven, 1996).

Seal whiskers detect water movements

How do pinnipeds orientate themselves under water? As most pinniped species feed in conditions under which visibility is drastically reduced, for example at night, at great depths or in murky waters, it has been particularly unclear how they succeed in finding food. Here we show that harbour seals (*Phoca vitulina*) can use their whiskers to detect minute water movements. The high sensitivity of this sensory system should allow a seal to gain hydrodynamic information resulting from movements of other aquatic animals, such as prey, predators or conspecifics.

In the wild, blind but well-nourished seals have often been observed¹, showing that these marine mammals can orientate themselves and hunt successfully without vision. As pinnipeds do not possess an active sonar system², the kind of sensory information used for prey detection and spatial orientation remained unknown until now. In the aquatic environment, one usable source of sensory information consists of water disturbances caused by moving organisms. Consequently, hydrodynamic sensory systems have evolved many times in aquatic animals³. But, although marine mammals are highly adapted to their aquatic environment, hydrodynamic sensory systems have

never before been described in this animal group.

The whiskers of pinnipeds have been proposed to be useful for detecting water movements⁴. These whiskers are heavily innervated by slowly and rapidly adapting afferent fibres that respond to tiny deflections of the hair shaft^{5–7}. Like rats⁸, seals and sea lions use their whiskers for active touch discrimination^{4,9}, and probably for obstacle avoidance¹⁰. However, it was unknown whether they can also use these tactile hairs for the detection of low-amplitude water movements, thus sensing distance from other objects.

Using a 'go/no-go' response paradigm, we determined detection thresholds of hydrodynamic stimuli (in the range 10–100 Hz) for a male harbour seal. Stimuli were generated by a sinusoidally oscillating sphere (diameter 10 cm, oscillation axis vertical) attached to a Ling vibrator. Electronic signals driving the vibrator were computer-generated (stimulus duration 3 s, rise and fall times 300 ms), converted from digital to analog and power-amplified.

We trained the seal to place its head in a hoop opposite to the sphere (Fig. 1a). Here, the animal pressed the tip of its lower jaw on a knob screwed on a jib that was welded to the lower edge of the hoop. In this way a defined distance (adjustable from 5 to 50 cm) between the tip of the whiskers and sphere was guaranteed (Fig. 1a). By knowing the exact distance between the whiskers and the sphere, we could calculate the effective stimulus amplitude¹¹. Optical and acoustic cues were excluded by placing removable eye caps and headphones (pink noise masking, 151 dB rel. 1 mPa) on the seal. The seal indicated the detection of a

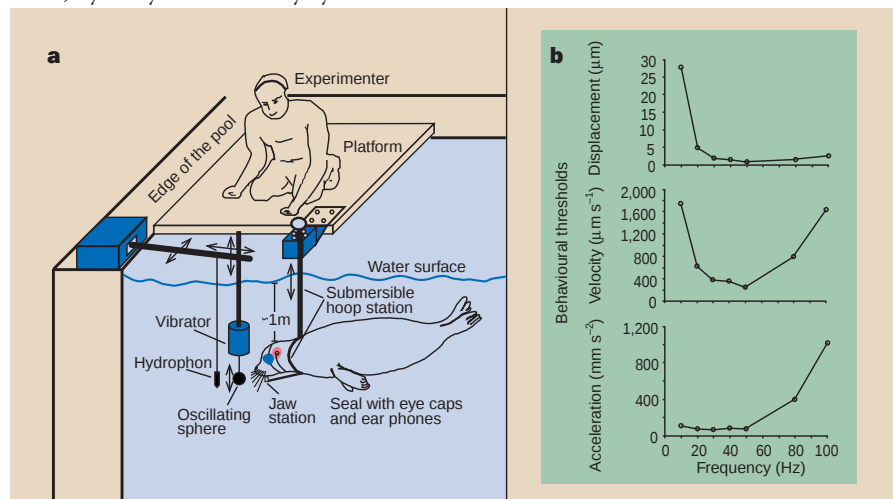


Figure 1 Detection of water movements by the harbour seal. **a**, Experimental set-up. At the beginning of a trial the hoop station was above the water surface, where the seal was stationed and supplied with eye caps and head phones, or, during whisker-exclusion tests, with a muzzle of wire mesh. Then the hoop was submerged with the seal to the final test position. The seal was trained to correct the position of its lower jaw whenever it lost contact with the knob of the jaw station. **b**, Behavioural displacement (top), velocity (centre) and acceleration (bottom) thresholds (defined as 50% correct decisions) of a harbour seal to sinusoidal water movements. Each threshold value is interpolated from 40 measurements with the last stimulus amplitude above threshold, and 40 measurements with the first stimulus amplitude below threshold.

water movement by leaving the hoop station during the period of stimulus presentation, after which it received a fish reward.

As soon as the seal was blindfolded, it brought its whiskers into the most forward position. Although the whiskers of seals can be easily moved by muscular action, as has been described for manatees¹², our test animal did not actively move the protracted whiskers during a trial.

The harbour seal responded to extremely weak hydrodynamic stimuli. However, it never responded when wearing a muzzle of wire mesh, which let the water movement through but impeded whisker movements. The shapes of the threshold curves (Fig. 1b) — plotted as particle displacement (μm), velocity ($\mu\text{m s}^{-1}$) or acceleration (mm s^{-2}) against frequency (Hz) of the stimulus — are similar to those of behavioural and physiological threshold curves for other aquatic animals (see ref. 3 and references therein). The nearly constant acceleration threshold values in the range 10–50 Hz indicate that at lower frequencies the seal responded to the acceleration component of the hydrodynamic stimulus. At higher frequencies, water displacement seems to be the relevant stimulus parameter. A similar change in operation mode is achieved by the fish lateral line³.

With regard to absolute thresholds, the sensitivity of the seal's whiskers to hydrodynamic stimuli compares well with that described for hydrodynamic receptor systems of, for example, the decapod crustacean *Cherax destructor* or the marine teleost *Xiphister atropurpureus*, whereas the lateral line system of some species of fish may be more sensitive by up to one or two orders of magnitude (see ref. 3 and references therein). The seal could detect water velocity at speeds as low as $245 \mu\text{m s}^{-1}$, which in turn is several orders of magnitude below the water-particle velocities measured in the wake of a swimming fish of about 22 cm in body length¹³.

Our results show that the whiskers of harbour seals form a hydrodynamic receptor system with a spectral sensitivity that is well tuned to the frequency range of fish-generated water movements¹³.

Guido Dehnhardt, Björn Mauck, Horst Bleckmann

University of Bonn, Institute of Zoology,
Poppelsdorfer Schloss, D-53115 Bonn, Germany
e-mail: dehnhardt@munster.de

8. Carvell, G. E. & Simons, D. J. *J. Neurosci.* **10**, 2638–2648 (1990).
9. Dehnhardt, G. & Dücker, G. *Anim. Learn. Behav.* **24**, 366–374 (1996).
10. Oliver, G. W. *Behaviour* **67**, 97–114 (1978).
11. Harris, G. G. & van Bergeijk, W. A. *J. Acoust. Soc. Am.* **34**, 1831–1841 (1962).
12. Reep, R. L. et al. *Mar. Mamm. Sci.* **14**, 257–273 (1998).
13. Bleckmann, H. et al. *J. Comp. Physiol. A* **168**, 749–757 (1991).

Alanine enantiomers in the Murchison meteorite

Engel and Macko¹ have reported that alanine indigenous to the Murchison meteorite has an L-enantiomer excess of about 33%. Furthermore, they reported very similar $\delta^{15}\text{N}$ values for L-alanine and D-alanine; these values are quite high in comparison with those of terrestrial amino acids. On this basis, they argued that contamination of the meteorite by terrestrial L-alanine can be ruled out because, if it contributed to the L-enantiomer excess, the lower ^{15}N content of terrestrial L-alanine would significantly lower the $\delta^{15}\text{N}$ value that they observed for L-alanine.

In agreement with earlier analyses², we find the alanine from the interior of Murchison to be racemic and believe that incomplete chromatographic resolution of L-alanine and other meteoritic amino acids may have affected both the enantiomer ratio and the $\delta^{15}\text{N}$ determinations of Engel and Macko.

The Murchison meteorite contains a complex suite of amino acids. Whereas terrestrial samples are dominated by the 20 protein amino acids, over 70 amino acids have been positively identified in this meteorite, many of which appear to be uniquely extraterrestrial. Others are present that have not been positively identified. This complexity challenges methods of amino-acid analysis. For example, in our study of enantiomeric ratios of meteoritic amino acids³ we found that we needed a fractionation

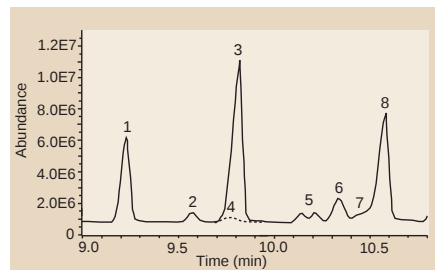


Figure 1 Total ion chromatogram (Chirasil-L-Val, 50 m x 0.25 mm, Alltech) of amino acids (*N*-trifluoroacetyl isopropyl esters) from a desalted Murchison water extract. The dotted line is a single ion (*m/z* 180) trace. Amino acids were identified from mass spectra. 1, D-alanine; 2, 2-amino-2-ethyl butanoic acid; 3, L-alanine; 4, methyl proline isomer or unsaturated acyclic C6 amino acid; 5, D-2-methyl norvaline; 6, D-*N*-methyl alanine; 7, unsaturated C4 amino acid; 8, sarcosine.

step before gas-chromatographic analysis in order to avoid resolution problems.

We find that several meteoritic amino acids are racemic, including alanine. An L-alanine excess was seen only in an extract of a sample that included exterior fragments. Figure 1 shows the chromatographic resolution of alanine from this extract, in the same chromatographic phase (Chirasil-L-valine) as that used by Engel and Macko¹. There are five amino acids that elute over a period of about a minute, from just before L-alanine to sarcosine. Although only one of these amino acids overlaps L-alanine in our chromatogram, in the separation shown in ref. 1 the chromatogram seems to be compressed and baseline separation is not quite achieved between L-alanine and sarcosine. Thus it seems probable that more than one of these five amino acids eluted or overlapped with L-alanine, contributing to the appearance of an L-alanine excess. It is unlikely that these amino acids were absent in Engel and Macko's sample as different Murchison stones are qualitatively similar with respect to amino-acid content⁴.

We do not know the $\delta^{15}\text{N}$ values of the amino acids that we suspect may be interfering but, if their $\delta^{15}\text{N}$ values were close to that of indigenous L-alanine, their presence could go undetected isotopically. On the other hand, if their $\delta^{15}\text{N}$ values were greater than that of L-alanine, their presence could mask the presence of terrestrial L-alanine. Given Murchison amino-acid values of up to +184‰, it is clear that their chromatographic overlap, even if they were present in relatively small amounts, would increase the ^{15}N content of the L-alanine peak and compensate, wholly or partially, for the lower ^{15}N content of terrestrial L-alanine that might have been present.

In conclusion, Engel and Macko's argument that their chromatographic peak for L-alanine is composed almost entirely of meteoritic L-alanine depends on the peak containing only L-alanine. Their argument for a large excess of the L-enantiomer in the alanine indigenous to their Murchison sample would be greatly strengthened if they were to provide evidence that L-alanine and the closely eluting amino acids that we have identified were completely resolved in their gas chromatography/isotope-ratio mass spectrometry analyses. Such analyses of appropriate amino-acid standards might also help to establish whether or not incomplete resolution affected their result.

S. Pizzarello, J. R. Cronin

Department of Chemistry and Biochemistry,
Arizona State University, Tempe,
Arizona 85287-1604, USA

1. Engel, M. H. & Macko, S. A. *Nature* **389**, 265–268 (1997).
2. Kvenvolden, K. et al. *Nature* **228**, 923–926 (1970).
3. Cronin, J. R. & Pizzarello, S. *Science* **275**, 951–955 (1997).
4. Cronin, J. R. & Pizzarello, S. *Adv. Space Res.* **3**, 5–18 (1983).

1. Newby, T. C. et al. *J. Mamm.* **51**, 152 (1970).
2. Watkins, W. A. & Wartzok, D. *Mar. Mamm. Sci.* **1**, 219–260 (1985).
3. Bleckmann, H. *Reception of Hydrodynamic Stimuli in Aquatic and Semiaquatic Animals* (Gustav Fischer, Stuttgart, Jena, New York, 1994).
4. Dehnhardt, G. & Kaminski, A. *J. Exp. Biol.* **198**, 2317–2323 (1995).
5. Hyvärinen, H. *J. Zool.* **218**, 663–678 (1989).
6. Dykes, R. W. *J. Neurophysiol.* **38**, 650–662 (1975).
7. Mills, F. H. J. & Renouf, D. *J. Exp. Mar. Biol. Ecol.* **100**, 3–9 (1986).

Microbiology for the millions

The Gospel of Germs: Men, Women, and the Microbe in American Life

by Nancy Tomes

Harvard University Press: 1998. Pp. 351.

£18.50, \$29.95

W.F. Bynum

For much of the twentieth century, historians accepted unquestioningly that the germ theory of disease was the single most important foundation of modern medicine, and that its application was directly responsible for the substantial improvements in life expectancy enjoyed throughout the developed world. In the 1960s, however, commentators such as the late Thomas McKeown, professor of social medicine at the University of Birmingham, began to argue that the old faith in 'germs' was historically naive, and that the decline of infectious diseases from the late nineteenth century owed much more to general social, economic and nutritional factors than to the specific activities of health workers. McKeown's work inspired a generation of historians for whom the claims of science, past and present, needed to be closely scrutinized, just like any other form of knowledge.

The McKeown thesis has now been substantially revised, and the final chapter of Nancy Tomes's fine study of the diffusion of germ theory into American popular culture provides an insight into why infectious agents have returned to centre stage: AIDS and other emerging diseases, and the spectre of drug resistance, have made us better able to appreciate our forefathers' fear of cholera or tuberculosis.

Tomes's reflections on the gospel of germs in the age of AIDS merely serve as an epilogue. The heart of her monograph focuses on the immediate legacy of Louis Pasteur, as the potentially dangerous world of microorganisms was revealed to the public. Pasteur's famous experiments on spontaneous generation had implicated particles of dust as an important source of contamination, and John Tyndall's researches in the 1870s reinforced the notion that the air we breathe could contain disease-causing germs. A writer in *The New York Times* had no doubt in 1874 that "an invisible cloud of scarlet-fever germs, typhoid seeds, and cholera or diphtheria spores" was present in the city, generated in overcrowded tenements but threatening the wealthy.

When Lucretia Garfield, wife of US President James Garfield, came down with a serious fever in the spring of 1881, the drains of the White House were suspected. When the president was shot a few months later, his doctors and the public despaired as to whether the presidential mansion was a fit place to nurse him. A sanitary engineer had

Sunbrite Cleanser
1-Cleans, Scours
2-Sweetens, Purifies
Double Action
Single Cost

KEEP YOUR STOVE SHINING!

You must know the easy way to polish the porcelain; to shine the metal on the stove. Sunbrite Cleanser does it in a twinkling—and dispels all the cooking odors at the same time. That's the secret of Sunbrite's unique "double-action". Sunbrite is so easy to use. The most stubborn stains vanish as soon as this capable cleanser gets to work. Let Sunbrite clean and purify your bathtub, refrigerator, utensils and sink. Saves you time and money. Saves your hands, too. Order a supply today.

Swift & Company

Quick Arrow Soap Flakes
Suds A Plenty

already been called in and had pronounced the plumbing defective and dangerous. As Garfield lingered, it was easy to blame the air he was breathing, and he was removed from the White House a few days before he died from his wounds. Congress belatedly authorized remedial work on the plumbing.

It might be tempting to see these and similar examples as nothing more than subtle modifications of traditional sanitarian notions. But this would be a mistake, as Tomes makes clear, because germ theory redefined the relationship between the environment and disease. It gave a new significance to 'dirt', and turned housekeeping into domestic science. Through the popular press, in books with such charming titles as *Household Bacteriology* and *Women, Plumbers and Doctors*, in courses in bacteriology and domestic science in schools and colleges, and through the activities of many local and national public-health organiza-

tions, the gospel of germs was brought to the public, and women were given a crucial role in the prevention of disease.

Tomes is particularly good with gender issues but she also shrewdly analyses race, particularly as it featured in the anti-tuberculosis campaigns of the early twentieth century. Germs were quickly commercialized, and vacuum cleaners, refrigerators and white porcelain sanitary fittings were marketed as essential accoutrements of the healthy home. Flies, fingers and food became major sources of disease and, when Du Pont developed Cellophane in the late 1920s, the company pointed out that now women could buy bread "untouched" by human hands. Listerine not only made the breath sweeter, it made kissing safer. Toilet tissue appeared just after the First World War, and drinking from communal cups was discouraged even earlier. Restaurants began to require their waiters to be clean-shaven, and

hotels were required to launder sheets after each occupant, and to use sheets sufficiently long that they could be folded back to protect against the potentially dangerous blanket.

In these and a thousand other ways, the gospel of germs influenced human behaviour and created our world. By the early twentieth century, scientific orthodoxy taught that people rather than things were the main receptacles of communicable disease, and household dust began to be seen as a nuisance rather than a death trap. Nevertheless, many household practices that started in the early decades of germ theory continued as established rituals, and polio replaced typhoid as an important source of parental anxiety. Polio vaccines and antibiotics seemed to render the gospel of germs an historical relic until the microbe reared its ugly head again in the past two decades.

Tomes's book is a fine example of the new social history of biomedical science, full of fascinating detail, elegantly written and cogently argued. In exposing the values and practices of our ancestors, it has much to teach us about ourselves. □

W. F. Bynum is at the Wellcome Institute for the History of Medicine, 183 Euston Road, London NW1 2BE, UK.

From meat to merchandise

Economy and Society in Prehistoric Europe: Changing Perspectives

by Andrew Sherratt
Edinburgh University Press/Princeton University Press: 1997. Pp. 561. £45, \$59.50

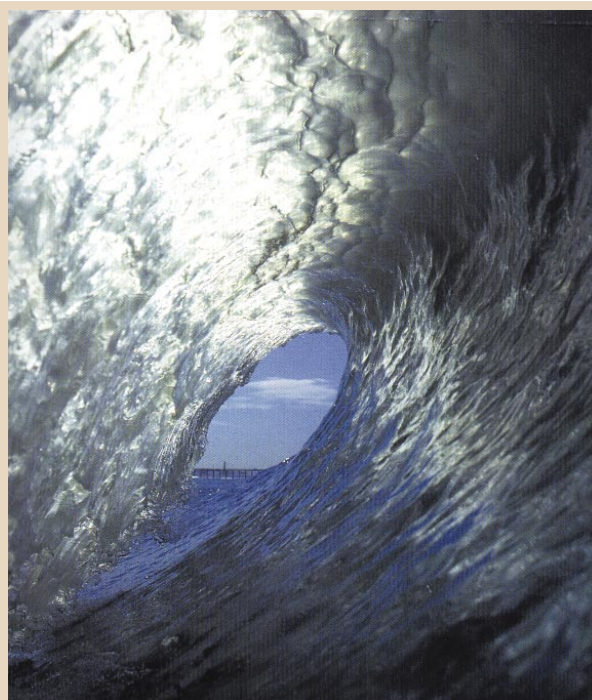
John Robb

Controversies about prehistoric social relations tend to be as opaque to outsiders as the theological quibbles in *The Name of the Rose*. Why does it matter whether Christ owned his sandals or merely used them? The answer, of course, is that such minutiae are the vocabulary of much broader polemics. In the same way, Andrew Sherratt's *Economy and Society in Prehistoric Europe* must be read on several levels. In building detailed factual arguments about wheeled vehicles, the plough or mind-altering substances, Sherratt is also constructing reference points in a particular skyline rising out of the urban sprawl of contemporary archaeological theory. It is for this reason that the essays collected in this book are likely to continue to arouse controversy.

Take the 'secondary products revolution', for example, which still remains Sherratt's best known Big Idea. This is the generalization that the earliest Eurasian farmers merely ate their animals; only several millennia later did they develop a suite of 'secondary products' for extracting more from livestock and

Surfing the wet

When you leave your computer or lab bench for a holiday by the sea, you may not care about the physics of wave motion, the origin of swell, or the fact that a water molecule has a circular motion and barely moves forwards at all as a wave passes. You may just want to go for a swim, watch waves break on the shore, or plunge headlong into the surf, board in hand. Either way, the third edition of Drew Kampion's *The Book of Waves: Form and Beauty on the Ocean* (Roberts Rinehart, \$29.95, pbk) shares your view. Part text, part photo gallery, it celebrates waves from ripples to breakers.



landscapes, such as dairying, wool production, seasonally migratory pastoralism, and traction devices such as the plough and cart. When first published in 1981, this concept changed the face of economic prehistory by punctuating the long, static stretch between the adoption of agriculture and the rise of Iron Age states, much as Colin Renfrew's introduction of the chiefdom concept did for social structure.

As an empirical generalization, the secondary products revolution has come in for its share of criticism. Many archaeologists would say that the evidence for the spread of ploughs, carts, dairying and wool use from the Near East, starting in the fourth millennium BC, is certainly suggestive but far from conclusive. Given the perishability of wood and other organic technologies, and the imprecision of economic reconstruction based on faunal remains, one source of nagging doubt remains the argument's reliance on the absence of early evidence for ploughing and pastoralism in regions such as western Europe. After all, many of the primary sources — rock art, sealings, writing and other administrative technology — are entirely lacking in much of Europe for all of prehistory. But it would take considerable expertise to argue these points with Sherratt over the breadth of European prehistory.

On the parallel, and elusive, plane of theory, archaeologists worry about how to fit images of the ancient world together in ways dictated only loosely, if at all, by facts. The concept of the secondary products revolution requires and creates a certain theoretical space that has evolved over the past couple of decades, as Sherratt notes in his introduction. He would now jettison the original idea — as much a hallmark of 1970s 'new archaeology'

as chiefdoms — that population pressure was an ultimate cause of the economic changes.

Thus freed, the concept meshes comfortably with the idea of a transition in the Late Neolithic and Early Bronze Age to an economy whose surplus production was used for prestigious exchange, display and consumption. Secondary products of cloth, food and drink were used to trade, to give away ceremonially, to use ostentatiously, and to feast. The combination of these two ideas forms a central plank of syntheses of European prehistory compiled in the past two decades.

But the theoretical landscape has broadened rapidly, and Sherratt's vision increasingly represents only a part of it, although one that is likely to endure. His focus on continent-scale patterns and coherent systems runs against currently popular attempts to explore the complexities and discontinuities of meaningful experience on a relentlessly local scale, and his models tend to give primacy to political and economic aspects of life over cultural identities. Exponents of many theoretical schools would discount his vision on these grounds at least as much as on the archaeological details. Yet Sherratt's own changing interests have generated new frames for the secondary products revolution picture — some fashionable (such as his focus on narcotics and intoxicants in prehistory), others stubbornly unfashionable (such as his redoubled efforts to trace the diffusion of new technologies from a Near Eastern cradle).

The works collected in this book present a wide range of Sherratt's thinking, including general discussions of prehistoric social structure, detailed models of Neolithic trade systems, intriguing attempts to trace the prehistory of alcohol and narcotics use, and

Sherratt's personal entries in the two traditional sweepstakes in later European prehistory: "What caused Neolithic megaliths?" and "Who were the Indo-Europeans?"

The entire volume stands as an attempt to live up to Sherratt's general conclusion: "To succeed in his or her trade, the prehistorian must always be a generalist." In Sherratt's case, the effort is clearly successful. The scenarios are imaginative and have impressive depth, and the evidence is clearly marshalled. Even when disagreeing with the approach or conclusions, one can learn a tremendous amount. □

John Robb is in the Department of Archaeology, University of Southampton, Southampton SO17 1BJ, UK.

No laughing matter

The Trembling Mountain: A Personal Account of Kuru, Cannibals and Mad Cow Disease

by Robert Klitzman

Plenum: 1998. Pp. 300. \$27.95, £16.94

Colin L. Masters

The outbreak of kuru among the Fore people of Papua New Guinea that climaxed in the mid-1950s has many potential lessons for contemporary UK society, as it tries to understand the emergence of a new variant of Creutzfeldt-Jakob disease (CJD), linked with the recent outbreak of bovine spongiform encephalopathy (BSE, or 'mad cow disease').

Robert Klitzman's breezy narrative

account of his fieldwork experiences in 1981 joins a growing list of monographs aimed at the wider community, notably *Fatal Protein: The Story of CJD, BSE and Other Prion Diseases* by Rosalind Ridley and Harry Baker (Oxford University Press, 1998). With varying degrees of success, these accounts convey the unprecedented effects of a devastating, infectious neurological illness, either in the setting of a Stone Age pre-scientific culture (the Fore) or in the technological sophistication of modern Europe (the British). With kuru, the Fore reacted with claims of sorcery, hysterical conversion reactions and retaliatory armed conflict. With the emergence of BSE and new-variant CJD, the British went through phases of denials, cover-ups and trade embargoes while the daily tabloid newspapers conducted their contemporary versions of hysterical reportage.

Klitzman's retelling of the kuru saga lacks the authoritative analysis one would have expected from an academic psychiatrist. But at the time of his first visit to Papua New Guinea, he was a naive Jewish boy from Long Island, New York, trying to decide on a future career in medicine. As a young college graduate, he had fallen under the spell of Carleton Gajdusek at the US National Institutes of Health, and travelled to Papua New Guinea to work with Michael Alpers in recording the disappearance of this infectious disease.

Although the transmissibility of kuru through cannibalism had been well established, there remained many unanswered questions, particularly relating to the mechanisms of transmission (route and dose of inoculum) and the effects of genetic suscep-



Seeking stability: in the language of the Fore people, the word *kuru* means 'to tremble'.

tibility. For example, were those who escaped infection participants at cannibalistic feasts at which they (the "natives"; Klitzman is very fond of this term) had eaten only the fingers and feet of the deceased? Klitzman searched and found examples of common point exposures with subsequent development of disease. He discovered the difficulties of doing fieldwork in kuru epidemiology where dirt roads were only just beginning to open up, most travel was on foot, and he was dependent on the cooperation of local guides whose senses of time and purpose were superficially different from those of a young New Yorker. ("The natives were skilled at getting what they wanted... They, too, were greedy, scheming, and deceptive when necessary.")

Themes of cultural diversity and evolution, the impact of western technology and the infusion of scientific endeavour into a rapidly changing Stone Age culture, which are clearly set out in the Gajdusek diaries and field notes (*Kuru: Early Letters and Field-Notes from the Collection of D. Carleton Gajdusek* by J. Farquhar and D. C. Gajdusek; Raven, 1981) fail to be conveyed by Klitzman's first-person narrative style. Although more compelling than the semi-fictional recollections of Vincent Zigas in *Laughing Death: The Untold Story of Kuru* (Humana, 1990), Klitzman's account is aimed at his fellow New Yorkers, most of whom have never heard of kuru, CJD or BSE, let alone Papua New Guinea.

There are surprisingly few first-hand accounts of the kuru story. During the next decade, when the full implications of the BSE crisis in Britain become known, it will be possible to look back on the emergence and

Bamboozled by habitat loss



Deforestation leaves giant pandas isolated and with nowhere to go when bamboo, their staple diet, periodically dies. Renowned photographer

Heather Angel visited China to capture these endangered animals on film. The results can be found in *Pandas* (Voyageur Press, \$16.95).

disappearance of kuru as an important episode in our understanding of the risks associated with this type of infectious process. Informing the wider community of these risks may lead to a more helpful debate about the public health policies required to minimize the chances of another BSE epidemic. Books such as this are useful in this context. □

Colin L. Masters is in the Department of Pathology, The University of Melbourne, Parkville, Victoria, 3052, Australia.

The making of a biochemist

Otto Warburgs Beitrag zur Atmungstheorie: Das Problem der Sauerstoffaktivierung*

by Petra Werner

Basilisk-Press: 1996. Pp. 390. DM136

Mikuláš Teich

The biochemist Otto Heinrich Warburg (1883–1970) was a name to conjure with in the days before molecular biology. In 1908 he began to investigate — in part with Otto Meyerhof — the respiratory activities of various materials, such as sea-urchin eggs, avian erythrocytes and liver tissue. His experimental results led him to surmise that cellular oxidation was linked to the catalytic activity of iron in cells.

After returning from the First World War, Warburg continued with his studies of respiration using cancerous tissue, and improved the manometric method of gas analysis. His starting point was that cellular respiration was a cyclic reaction. Oxygen combined with iron to produce higher-valency iron, which reacted with oxidizable organic material and in so doing returned to the bivalent state. Warburg might have obtained the idea of the cycle from Meyerhof, who in 1919 and 1920 studied chemical changes occurring in muscle in relation to the work done or the energy liberated as heat. Meyerhof visualized the breakdown of carbohydrate to lactic acid as the anaerobic phase, and the synthesis of carbohydrate as the aerobic phase, of a ‘specific carbohydrate cycle’.

Warburg’s investigations of artificial iron-containing systems, which were presumed to be analogous to reactions occurring in living cells, satisfied him that iron functions as the oxygen-transferring constituent of a ubiquitous cellular catalyst, which he named *Atmungsferment* (‘respiratory ferment’). He was a most inventive experimenter. Among the artificial systems that he studied were the oxidation of amino acids by haemin charcoal, and the inhibition of oxygen uptake by cyanide and urethane.

* Otto Warburg’s Contribution to Respiration Theory: The Problem of Oxygen Activation.

In the late 1920s, he looked into the effect of light on the inhibition by carbon monoxide of respiration in living cells. This work encompassed considerations of photochemical processes in terms of quantum chemistry, and the use of the manometer, photoelectric cell and spectroscope. From the shape of the curve obtained by plotting the effectiveness of light against its wavelength, it was possible to deduce the resemblance between the respiratory ferment and haemins. Warburg was awarded the Nobel prize for physiology or medicine in 1931 for his recognition of the haemin-type nature of the respiratory ferment and its underlying principles.

The development of Warburg’s theoretical thinking and experimental procedures are ably chronicled in Petra Werner’s introductory essay. Her book is the first volume of an edition of Warburg’s correspondence deposited in the Berlin–Brandenburg Academy of Sciences. Regrettably, the 143 published letters covering the period 1906 to 1939 include only 14 by Warburg, all of which were to Jacques Loeb. Warburg’s early work was strongly influenced by Loeb’s book, published in 1906, which dealt with artificial parthenogenesis and the nature of fertilization in thoroughly physicochemical, reductionist terms. Loeb immediately recognized in Warburg a kindred spirit, and was prepared to get him a grant from the Rockefeller Institute to work in his laboratory, and to help him settle in the United States. Before and after the First World War, Loeb — an early emigré from Germany as a result of his experiences of anti-Semitism in the academic world — befriended not only Warburg, but also Meyerhof and Leonor Michaelis, who contributed to the development of a mathematical theory of enzyme processes. This emerges from Loeb’s letters to the two scientists, which are also included in the book.

Outstanding as Warburg was as a scientist, even his admiring Nobel prize-winning research students, Hans Krebs and Hugo Theorell, realized that he tended to pettiness. There can be little doubt that this fuelled his resentment of Meyerhof, who, with Archibald Vivian Hill, won the Nobel prize for physiology or medicine nine years before Warburg for work on muscle metabolism. Werner also refers to Warburg’s selective approach to history in two retrospective monographs: “he cited only Nobel prize winners or widely known persons... and left others out; thus, later he never mentioned the name of Jacques Loeb. This retrospective account submerged historical reality beneath an embellished, teleological presentation.”

The English version of Warburg’s first publication appeared as *Heavy Metal Prosthetic Groups and Enzyme Action* (Oxford University Press, 1949) and was critically reviewed by David Keilin (*Nature* 165, 4–5;



Brilliant but flawed: Warburg tended to pettiness.

1950). In 1925, Keilin demonstrated the reversible oxidation and reduction of a pigment, which he named ‘cytochrome’, in the thoracic muscles of the adult fly *Gasterophilus intestinalis*. This was a crucial event in the history of biological oxidation and helped lead to the interpretation of cell respiration in terms of a sequence of reactions driven by oxidation and reduction (the ‘respiratory chain’). That Keilin’s achievement did not earn him a Nobel prize was, I think, a notable omission. It is unfortunate that his name crops up just once in the book.

That said, the book is an important source for studying the development of the chemistry of life from about 1900 to 1930. □ Mikuláš Teich is at Robinson College, University of Cambridge, Cambridge CB3 9AN, UK.

Software reviews at www.nature.com

From this week, *Nature*’s website presents comparative reviews of scientific software.

Nature has recruited a group of reviewers to test a wide range of scientific software, including graph-making and statistics packages, mathematics software, systems for bibliography and reference management, and more. This week, Sharon Kardia of the University of Michigan inaugurates the series at www.nature.com, with reviews of 16 graph-making packages.

Kardia explains the criteria used to evaluate the packages, and outlines the strengths and weaknesses of each. She calls attention to unique features, and suggests who might benefit from using each package. A table of system functions shows which packages possess which capabilities, and gives performance ratings for each package, for each of these functions. Each review is hyperlinked to additional product details provided by the manufacturer.

High-redshift star formation in the Hubble Deep Field revealed by a submillimetre-wavelength survey

David H. Hughes*, Stephen Serjeant†, James Dunlop*, Michael Rowan-Robinson†, Andrew Blain‡, Robert G. Mann†, Rob Ivison*, John Peacock*, Andreas Efstathiou†, Walter Gear§, Seb Oliver†, Andy Lawrence*, Malcolm Longair‡, Pippa Goldschmidt† & Tim Jenness||

* Institute for Astronomy, University of Edinburgh, Royal Observatory, Edinburgh EH9 3HJ, UK

† Astrophysics Group, Imperial College, Blackett Laboratory, Prince Consort Road, London SW7 2BZ, UK

‡ Cavendish Astrophysics Group, Cavendish Laboratory, Madingley Road, Cambridge CB3 0HE, UK

§ Mullard Space Science Laboratory, University College London, Holmbury St Mary, Surrey RH5 6NT, UK

|| Joint Astronomy Centre, 660 N. A'ohoku Place, Hilo, Hawaii 96720, USA

In the local Universe, most galaxies are dominated by stars, with less than ten per cent of their visible mass in the form of gas. Determining when most of these stars formed is one of the central issues of observational cosmology. Optical and ultraviolet observations of high-redshift galaxies (particularly those in the Hubble Deep Field) have been interpreted as indicating that the peak of star formation occurred between redshifts of 1 and 1.5. But it is known that star formation takes place in dense clouds, and is often hidden at optical wavelengths because of extinction by dust in the clouds. Here we report a deep submillimetre-wavelength survey of the Hubble Deep Field; these wavelengths trace directly the emission from dust that has been warmed by massive star-formation activity. The combined radiation of the five most significant detections accounts for 30–50 per cent of the previously unresolved background emission in this area. Four of these sources appear to be galaxies in the redshift range $2 < z < 4$, which, assuming these objects have properties comparable to local dust-enshrouded starburst galaxies, implies a star-formation rate during that period about a factor of five higher than that inferred from the optical and ultraviolet observations.

Recent years have seen the first meaningful attempts to determine the global star-formation history of the Universe, using the combined information provided by deep redshift surveys (for example, the Canada France Redshift Survey¹) reaching $z \approx 1$, and the statistics of Lyman-limit galaxies² at higher redshifts in, for example, the Hubble Deep Field (HDF)^{3–5}. The results⁶ imply that the star-formation and metal-production rates were about 10 times greater at $z \approx 1$ than in the local Universe, that they peaked at a redshift in the range $z \approx 1–1.5$, and that they declined to values comparable to those observed at the present day at $z \approx 4$.

These conclusions, derived from optical-ultraviolet data, may however be misleading, because the absorbing effects of dust within distant galaxies undergoing massive star-formation may have distorted our picture of the evolution of the high-redshift Universe in two ways. First, the star-formation rate (SFR) in known high-redshift objects is inevitably underestimated unless some correction for dust obscuration is included in deriving the rest-frame ultraviolet luminosity. Second, it is possible that an entire population of heavily dust-enshrouded high-redshift objects, as expected in some models of elliptical galaxy formation⁷, have gone undetected in the optical-ultraviolet surveys. The extent of the former remains controversial^{8–11}, while the possibility of the latter has until now been impossible to investigate.

Submillimetre cosmology

At high redshifts ($z > 1$), the strongly-peaked far-infrared radiation emitted by star-formation regions in distant galaxies is redshifted into the submillimetre waveband, and the steep spectral index of this emission on the long-wavelength side of the peak, at $\lambda \approx 100 \mu\text{m}$ in the rest-frame, results in a rapid increase in the rest-frame luminosity probed by submillimetre observations with increasing redshift, sufficient to offset the dimming of galaxies due to increasing cosmological distance. Consequently the flux density

of a galaxy at $\lambda \approx 800 \mu\text{m}$ with fixed intrinsic far-infrared luminosity is expected to be roughly constant at all redshifts in the range $1 \leq z \leq 10$ (refs 12–14).

This ease of access to the young Universe has already been exploited through successful pointed submillimetre observations of known high-redshift sources including lensed objects (IRAS F10214+4724¹⁵ and the Cloverleaf quasar¹⁶), radio galaxies^{14,17,18} and quasars^{19,20}. These studies have demonstrated the potential of submillimetre cosmology and have shown that in at least some high-redshift galaxies, dust-enshrouded star formation is proceeding at a rate of $\gg 100 M_{\odot} \text{yr}^{-1}$, substantially greater than the more modest star-formation rates (on average $\sim (1–5)h^{-2} M_{\odot} \text{yr}^{-1}$) displayed by Lyman-break galaxies³. (Here $M_{\odot}$ is the solar mass, and h is the normalized Hubble constant in units of $100 \text{ km s}^{-1} \text{Mpc}^{-1}$).

With the recent commissioning of the sensitive submillimetre array camera SCUBA on the James Clerk Maxwell Telescope (JCMT)²¹ it is now possible to conduct unbiased surveys²² in this wavelength region and quantify the amount of star-formation activity in the young Universe by observing directly the rest-frame far-infrared emission from dust in high-redshift galaxies. Here we describe the first results from an ultra-deep submillimetre survey centred on the HDF.

A submillimetre survey of hidden star formation in the HDF

Recent ISOCAM observations of the HDF at 6.7 and $15 \mu\text{m}$ have confirmed that the strong evolution seen in the IRAS galaxy population at low redshifts^{23,24} continues out to redshifts of the order of unity^{25,26}. Such mid-infrared studies cannot, however, provide any constraints at higher redshift. In contrast an 850- μm survey is predicted to be completely dominated by sources at $z \geq 1$, and the number of detectable sources is very sensitive to the high-redshift evolution of the dusty starburst population. In particular, a SCUBA survey of the HDF complete to a flux density limit

$S_{850\mu\text{m}} > 2 \text{ mJy}$ would be expected to detect <0.1 galaxies if there is no cosmic evolution, <1 galaxy if the evolution mirrored the Madau curve⁴, but at least 2 sources if the number density and luminosity of infrared starburst galaxies continued to evolve strongly out to $z \approx 2$, and substantially more sources with $z > 2$ if the population continued to evolve or stayed constant at higher redshifts^{7,12,27–29}.

We chose to centre this deep 850- μm survey on the HDF, not only because the SCUBA field of view of $\sim 6 \text{ arcmin}^2$ is well matched to the area of the HDF, but also to maximize the possibility of finding optical/infrared/radio counterparts and redshifts for any submillimetre sources which were detected. Currently there exist over 20 spectroscopic redshifts for galaxies at $z > 2$ within the HDF, while the availability of deep photometric data³⁰ in the U_{300} , B_{450} , V_{606} and I_{814} bands facilitates the estimation of photometric redshifts for other galaxies in the field.

Simultaneous diffraction-limited images of the HDF at 850 and 450 μm were taken with SCUBA²¹ on the 15-m JCMT. A total of 50 hours integration between 5 January and 13 February 1998 were centred at right ascension (RA) 12 h 36 min 51.2 s, declination (dec.) $62^\circ 12' 52.5''$ (J2000) with occasional offsets 25 arcsec south, east and west to aid the discrimination of real and spurious sources. The data were taken under exceptional atmospheric conditions, with a median 850- μm sky opacity $\tau_{850\mu\text{m}} = 0.16$. Sky subtraction was performed using on-array chopping in RA in order to minimize the chop throw (important for accurate sky subtraction), to maximize the reclaimable signal-to-noise ratio for detected sources, and to minimize the number of negative off-beams arising from unknown sources well outside the primary field of view. We experimented with chopping in azimuth, but, at least at the declination of the HDF, this yielded no significant noise improvement over chopping in RA. Finally, to ensure that no significant source would be missed due to an unfortunate coincidence with the off-beam of another brighter source, the length of chop throw was varied, approximately half the observations (29 hours) adopting an RA chop-throw of 30 arcsec, and the remainder (21 hours) an RA chop-throw of 45 arcsec. As discussed below, this approach proved invaluable both for source confirmation, and for the separation of real and confused sources. The 850- μm data, with an angular resolution of 14.7 arcsec full-width at half-maximum (FWHM), covers an area of $\sim 9 \text{ arcmin}^2$ and, due to the variation in the density of bolometer samples across the map, has a noise at the periphery approximately double its value at the map centre. The 850- μm image in Fig. 1 shows a circular field, within a radius of 100 arcsec from the map centre, and reaches a 1σ noise level of 0.45 mJy per beam. This image represents the deepest submillimetre map taken.

Submillimetre source extraction and confusion

Because the noise increases with radius from the map centre, sources were only sought within the central 80-arcsec radius of the image. The map shown in Fig. 1 displays 58 distinct peaks, most of which are noise. For gaussian filtered white noise, 1% of the peaks exceed 3.3σ in amplitude³¹, and so a flux density of 1.5 mJy (at the map centre) is the practical detection threshold for real sources; the map contains 7 such objects. The use of two different chops and the effect of telescope nodding is to produce a convolving beam with four negative sidelobes. The signature of a source is therefore very different from noise, and this fact can be used to identify real sources and to deconvolve the map down to some flux density limit. Deconvolution also allows the flux in the sidelobes of a source to be reclaimed, thereby enhancing the signal-to-noise ratio of the detected sources. To investigate whether these peaks correspond to single, or blended, sources, simulations of random source distributions with plausible number counts have been performed. The 14.7-arcsec beam is sufficiently broad that an ideal noise-free map would in fact never contain more than ~ 20 peaks within the 5.6 arcmin^2 map, independent of the true density of sources. Alternatively, the

observed source density is about one source per 12 beam areas; both arguments indicate that source confusion must become important at the limit of our 850- μm map.

It is thus possible that at least some of the apparent sources in the map could consist of emission from more than one object, and this is a particular concern for the weaker sources with $S_{850\mu\text{m}} \approx 2 \text{ mJy}$. One way of isolating such cases is optical identifications, as discussed below; if there is only a single candidate identification, the source cannot be a blend, as each member of the blend would have a separate optical counterpart.

Another approach is to note that confusion is only a serious problem when there is a blend of one or more sources of similar flux, and that in such cases the apparent source will usually be significantly broader than the telescope beam. This breadth means that the apparent source position will be less stable under the addition of noise than if the source is dominated by a single unresolved object. We have therefore taken the conservative approach of identifying sources in the full data set that appear in both the 30-arcsec and 45-arcsec chop images, and only keeping those whose positions agree to better than 3 arcsec. Tests on simulated source fields with realistic number counts show that this procedure should succeed in giving a clean sample of the sources brighter than 2 mJy in the central 80-arcsec radius of the image, each dominated by a single object. The positions of these 5 sources are given in Table 1, together with their 850- μm flux densities.

The simultaneous 450- μm image covers 75% of the useful area mapped at 850 μm and, despite the excellent observing conditions, which resulted in a 1σ r.m.s. noise signal of 7 mJy per beam at 450 μm , no significant detections were obtained.

Number counts and the submillimetre background

The map shown in Fig. 1 can be used to determine the form of the number counts at 850 μm fainter than the limit of 2 mJy at which

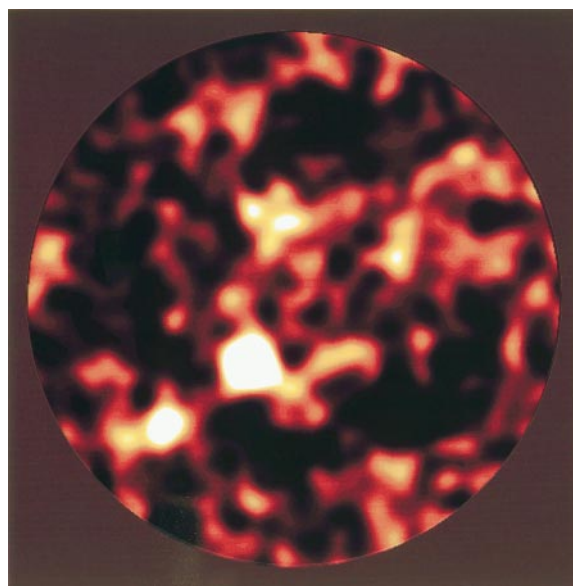


Figure 1 The 850- μm SCUBA image of the HDF. The image shows a radius of 100 arcsec from the map centre (RA 12 h 36 min 51.20 s, dec. $+62^\circ 12' 52.5''$; J2000) and is orientated with north upwards and east to the right. Primary flux calibration was performed using Uranus, with secondary calibration against a variety of AGB stars and compact H II regions⁵⁴. A signal-noise image is shown, allowing the significance of faint potential sources to be judged in a uniform manner, although it means there is a tendency for more sources to be detected in the central regions. The colour coding has been tuned so that the 5 most significant sources discussed here appear in white, while other formally significant flux-density peaks appear yellow. The absolute calibration uncertainty is $<10\%$. See Methods for details.

individual sources can be selected with some confidence. Fainter sources combine to raise the r.m.s. fluctuations in the map beyond what is expected purely from noise. There is a long tradition in radio and X-ray astronomy of extracting faint counts from such information using ‘ $P(D)$ ’ analyses³², although the present data set is unusual in that both random noise and confusion noise are of similar amplitude. The approach adopted here is to focus on the distribution of signal-to-noise ratios for the peaks of the map in Fig. 1 (that is, the distribution of fluxes for all apparent ‘sources’). By generating synthetic maps with different number counts, it is possible to estimate what range of true counts is consistent with the observed distribution.

We have not explored the full parameter space, but some examples are illustrated in Fig. 2. Empirically, it is clear that there is an excess of peaks in the range 0.8–1.5 mJy, and this requires a substantial density of sources at about this flux-density level. The observed peak flux-density distribution is matched reasonably well by a source density of about 7,000 deg^{−2} brighter than 1 mJy, which corresponds to the observed density of brighter sources, extrapolated with a euclidean count slope, with the significant caveat that this number assumes an unclustered source distribution. If in fact the faint submillimetre sources are high-redshift starbursts, it is not implausible that they are strongly clustered on scales of several arcsec^{33,34}. For a given surface density of sources, this increases the background fluctuations and so the above figure should probably be treated as an upper limit.

The counts must continue to flux densities somewhat fainter than 1 mJy, but the present data do not have the sensitivity to estimate where the inevitable break from the euclidean slope occurs. This is best constrained by asking at what flux density the extrapolated count exceeds the background. By summing the flux densities in Table 1, a lower limit to the background contributed by discrete sources of 20 mJy per 5.6 arcmin² is found, equivalent to $\nu I_\nu = 1.5 \times 10^{-10} \text{ W m}^{-2} \text{ sr}^{-1}$, or approximately half the background estimate reported by Puget *et al.*³⁵ (where ν is the observing frequency and I_ν the specific intensity of the observed radiation at that frequency). There is, however, evidence in our data—specifically by continuing the deconvolution until the residual noise is statistically symmetric, or using the cumulative counts to 1 mJy derived above—that the true background contributed by discrete sources may be up to a factor of two higher than this, essentially

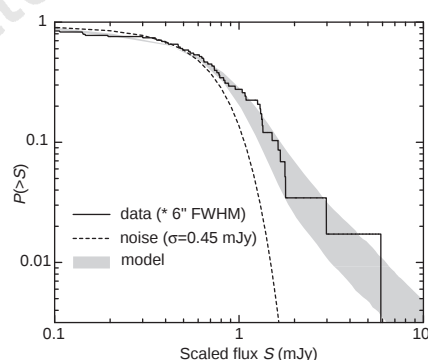


Figure 2 A comparison of observed and predicted number counts of submillimetre sources in the HDF as a function of flux density. The jagged solid line shows the raw integral number counts for the central 80-arcsec radius of the map in Fig. 1. The flux units here are signal-to-noise, but scaled to the flux units in the map centre. This uniform-noise representation allows a clear demonstration of sources in excess of the expectation for a pure noise field (dashed line) above ~0.8 mJy. Synthetic maps were made with the observed noise properties using a random distribution of sources having euclidean counts down to a limit of 0.3 mJy (although the results are insensitive to this cut-off). The grey band shows the effect of varying the integral surface density at 1 mJy between 4,000 and 10,000 degree^{−2}.

identical to the original estimate of Puget *et al.*, and consistent with more than 50% of the revised background estimates at 850 μm (refs 36, 37) which suggest $\nu I_\nu = (5.0 \pm 4) \times 10^{-10} \text{ W m}^{-2} \text{ sr}^{-1}$. The faint counts must therefore flatten by a flux density of ~0.3 mJy, otherwise even this background estimate would be exceeded.

Photometric observations and redshift estimation

Additional photometric observations during February 1998 at the centroid position of HDF850.1 confirmed the detection of the brightest submillimetre source with detections at 1,350 μm of $2.1 \pm 0.5 \text{ mJy}$ and at 850 μm of $7.0 \pm 0.4 \text{ mJy}$. These data, together with a 450- μm 3σ upper limit of 21 mJy per beam, provide a robust photometric estimate of the redshift of the source. In Fig. 3 the expected flux density ratios at submillimetre and millimetre wavelengths are plotted as a function of redshift for a range of models, typical of dusty, star-forming galaxies, which are consistent with the observed optical to submillimetre spectra of Arp220, one of the most heavily enshrouded local starburst galaxies^{38,39}, and the high- z starburst/active galactic nucleus IRAS F10214+4724. The relevance of these models to galaxies in the high- z Universe is reinforced by noting that the measured submillimetre and millimetre wavelength flux-density ratios of known high- z objects^{14,40} lie close to, or within the bounds of, the models.

The photometric redshift for HDF850.1, determined from the 1,350/850 μm flux-density ratio, lies within the range $2.5 < z < 9$.

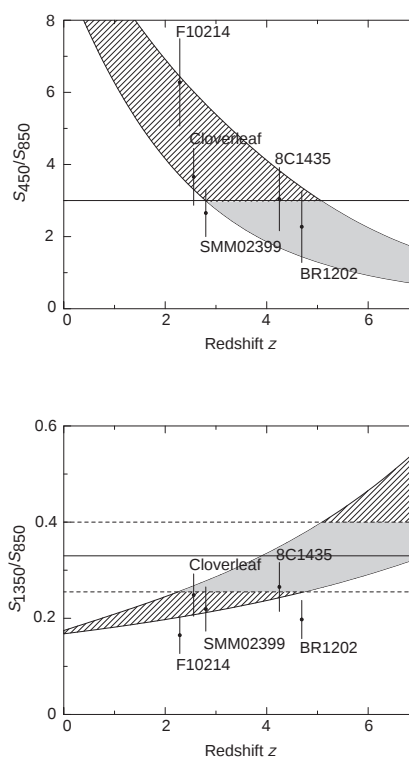


Figure 3 An estimation of redshift using the measured flux density ratio between wavelengths of 450 and 850 μm (top panel) and 1,350 and 850 μm (bottom panel). The hatched area in each plot shows the expected range of flux density ratio as a function of redshift bounded by two extreme models of dusty star-forming galaxies³⁸ (Arp220 and IRAS F10214+4724) which are consistent with observations of high- z galaxies^{12,40}. The solid horizontal lines represent the measured flux ratios for HDF850.1 and, in the case of the $S(1,350 \mu\text{m}/850 \mu\text{m})$ ratio, the horizontal dotted lines represent errors of $\pm 1\sigma$ on the observed ratio. The solid shading represents the parameter space satisfied by the photometric data for HDF850.1, including the non-detection at 450 μm , and illustrates that the redshift for HDF850.1 probably lies in the range $2.5 < z < 9$.

This strong constraint is supported by its non-detection at 450 μm which provides an upper limit to the 450/850 μm flux-density ratio and hence a lower limit to its redshift of $z > 3$. Less stringent, but similar, high-redshift limits can be estimated for all submillimetre sources detected in the HDF by arguing that their non-detection at 15 μm at a 3σ level of $\sim 20 \mu\text{Jy}$ (S. Oliver *et al.*, manuscript in preparation) implies a lower limit of $z \approx 2$ for sources at 850 μm brighter than 2 mJy, assuming a starburst galaxy model³⁸, or $z > 1.5$ assuming the observed spectral energy distribution (SED) of the extreme starburst galaxy Arp220. The radio far-infrared correlation⁴¹ yields lower limits of $z = 1.75$ and $z = 2.75$ respectively for 2 mJy and 7 mJy sources detected at 850 μm , but not detected at 8.5 GHz (ref. 42) at a 5σ flux limit of 8 μJy . The above data demonstrate that deep submillimetre surveys provide an efficient means of identifying a population of star-forming galaxies at redshifts > 2 .

Optical associations with submillimetre HDF sources

Given the rest-frame optical/far-infrared ratios typical of luminous starburst galaxies, the high-redshift SCUBA-selected galaxies are not necessarily expected to be present in the optical HDF, despite its depth. Nevertheless, we briefly discuss in turn plausible associations

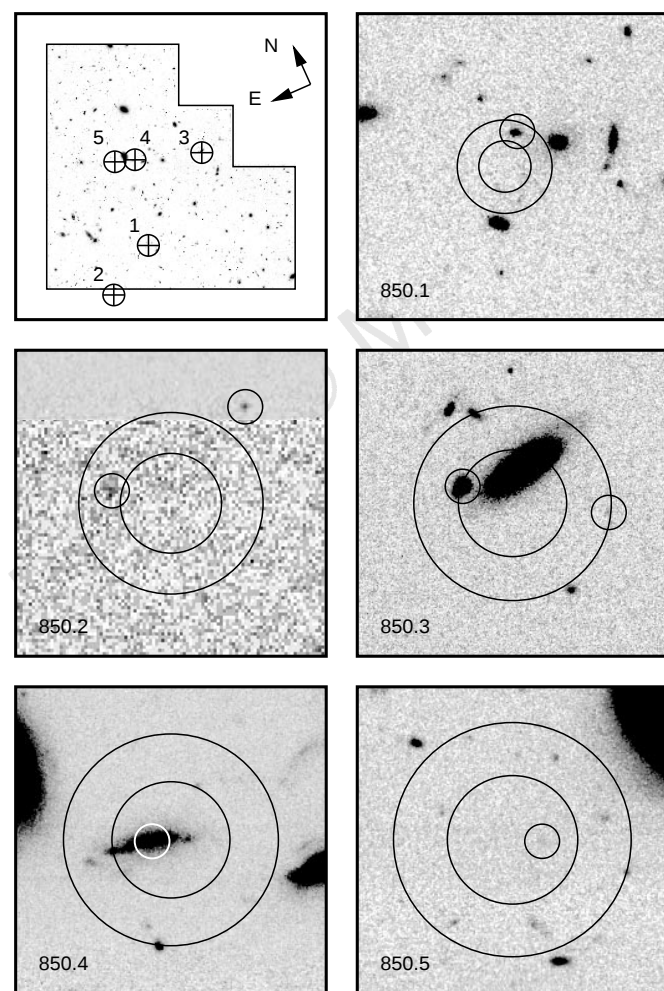


Figure 4 Optical associations for the brightest five submillimetre sources in the HDF. The top-left panel indicates the approximate location and orientation of each of the 10×10 arcsec/ $_{814}$ -band images shown in the following five panels. In each of these panels, the two large circles represent the 90% and 50% confidence limits on the submillimetre positional uncertainty, whilst a small circle (1 arcsec in diameter) has been used to mark the location of the most plausible optical association (or associations) for each SCUBA source (see text).

for the five most secure submillimetre sources in the HDF (see Fig. 4), estimating photometric redshifts⁴³, z_{ph} , for those galaxies without spectroscopic redshifts extended to include limits where galaxies are not detected in all four HDF bands.

Our approach is as follows. For each SCUBA source, we have considered as a potential optical counterpart all galaxies detected in the HDF whose distance from the SCUBA source lies within the 90% confidence limit of the submillimetre source position listed in Table 1. For each candidate we have then calculated the probability that a galaxy with such an optical magnitude (or brighter) could lie so close to the SCUBA position by chance, and also the probability that a galaxy with the observed redshift (or higher) could lie so close to the SCUBA position by chance. We note that these probabilities are often substantially higher than the raw Poisson probabilities⁴⁴. This is due to the combined effect of the rather large uncertainty in the SCUBA positions, and the high surface density of galaxies at the limit of the optical HDF image, which together essentially guarantee that (with the exception of HDF850.1) every SCUBA source will have at least one optical identification candidate at the limit of the HDF image. Finally we have investigated whether any of the apparently most probable optical identifications can in fact be clearly rejected on the basis of the SED constraints discussed above.

HDF850.1. As shown in Fig. 4, this source lies 1.0 arcsec from galaxy 3–577.0 in the optical HDF catalogue³⁰ which has a tentative spectroscopic redshift of $z = 3.36$ (ref. 45), which has been claimed⁴⁶ to be gravitationally lensed by the foreground $I_{814}(AB) = 24$ elliptical galaxy 3–586.0 which lies at $1.0 \leq z \leq 1.2$ (refs 43, 47, 48) (where AB magnitude is as defined by Oke⁴⁹). More recently, a 3.5σ detection (6.3 μJy) at 8.5 GHz has been associated with 3–577.0 (ref. 42). Based on its magnitude, the probability that 3–577.0 is a chance association with HDF850.1 is 0.33, while based on its redshift (which we estimate is $z_{\text{ph}} = 3.1$), the probability (calculated from the surface density of $z_{\text{ph}} > 3$ galaxies in photometric redshift catalogues^{43,47,48}) is only 0.20. For 3–586.0 the probabilities are in fact comparable (0.29 and 0.49 respectively), but the non-detection of HDF850.1 at 15 μm ($S(3\sigma) < 23 \mu\text{Jy}$) is strongly inconsistent (by almost two orders of magnitude) with the observed SEDs of any known galaxy (including Arp220) if placed at the ‘low’ redshift of 3–586.0, and as discussed above, the millimetre/submillimetre flux ratios also indicate that $z > 2.5$. Figure 5 shows that the observed spectrum of HDF850.1 agrees well with that expected for a starburst galaxy at redshift $z \approx 3$. We note that the random probability of being 2 arcsec from one of the radio sources⁴² is only 0.03. If the radio source really is associated with 3–586.0, and HDF850.1 with 3–577.0, then these seemingly incompatible probabilities are best explained by assuming that 3–577.0 is indeed being gravitationally lensed by 3–586.0, thereby amplifying its rest-frame far-infrared flux, and increasing its chances of being detected at 850 μm by SCUBA. The amplification would, however, need to be fairly substantial to explain the statistics, implying a massive lens:

Table 1 Positions and flux densities for the five most reliable submillimetre sources in the HDF

Source	IAU name	RA (h m s)	Dec. ($^{\circ}$ ' '')	$S_{850 \mu\text{m}}$ (mJy)
HDF850.1	J123652.3+621226	12 36 52.32 (± 0.10)	+62 12 26.3 (± 0.7)	7.0 ± 0.5
HDF850.2	J123656.7+621204	12 36 56.68 (± 0.20)	+62 12 03.8 (± 1.4)	3.8 ± 0.7
HDF850.3	J123644.8+621304	12 36 44.75 (± 0.21)	+62 13 03.7 (± 1.5)	3.0 ± 0.6
HDF850.4	J123650.4+621316	12 36 50.37 (± 0.23)	+62 13 15.9 (± 1.6)	2.3 ± 0.5
HDF850.5	J123652.0+621319	12 36 51.98 (± 0.25)	+62 13 19.2 (± 1.8)	2.1 ± 0.5

All these sources have $S_{850} > 2$ mJy. The positions of these sources (J2000) are reproduced from deconvolutions of both the 30-arcsec and 45-arcsec chopped images to within 3 arcsec. Positions given for each source were obtained from the average of the positions obtained using the independent SURF and IDL reductions. The quoted r.m.s. positional uncertainties were derived from the formula $\sigma_{\text{pos}} = \theta_{\text{beam}} / (2S/N)$, where θ_{beam} is the FWHM of the beam and S/N is the signal-to-noise ratio of the source. A further 0.5-arcsec uncertainty was added in quadrature, to account for the standard error in absolute pointing derived from measured pointing offsets throughout the observations. Flux densities quoted are the average from 3 independent methods, all of which agreed to within the formal uncertainty. Absolute calibration is uncertain to 10%.

3–586.0 may be the only visible member of a fainter group of galaxies.

HDF850.2 lies just beyond the edge of the HDF, making an assessment of possible optical associations difficult because the I_{814} band Hubble Flanking Field (HFF) image only reached a depth of ~ 25 mag. HDF850.2, however, is 4.3 arcsec from the $z_{\text{ph}} = 3.8$ galaxy 3–962.0 on the edge of the HDF, but based on its magnitude, the probability that 3–962.0 is a chance association with HDF850.2 is 0.63, while based on its redshift, the probability is 0.46. As can be seen in Fig. 4, there does appear to be a more convincing, but also very faint, candidate identification within the HFF, but at present we possess little useful colour information for this object. Therefore, although the non-detection at $15\text{ }\mu\text{m}$ implies a flux density ratio $S(850\text{ }\mu\text{m}/15\text{ }\mu\text{m}) > 190$ consistent with the SED of a starburst galaxy at $z > 2$, we are unable to make an unambiguous optical association.

HDF850.3 lies only 1.3 arcsec from 1–34.2 which is an asymmetric galaxy with $I_{814}(AB) = 24.5$ for which we estimate $z_{\text{ph}} \approx 1.95$. This is a moderately convincing identification, because, based on its magnitude, the probability that 1–34.2 is a chance association with HDF850.3 is only 0.29 (although based on its redshift the probability is 0.52) and its estimated redshift is consistent with a non-detection or marginal detection at $15\text{ }\mu\text{m}$. The next nearest object is 1–34.0, a $I_{814}(AB) = 21$ galaxy at a distance of only 1.5 arcsec. For this galaxy, the random probabilities are 0.12 and 0.60, respectively, but with a tentative spectroscopic redshift of 0.49, and photometric redshift estimates in the range $0.26 \leq z_{\text{ph}} \leq 0.68$, this object can be confidently rejected as a possible identification given its non-detection at $15\text{ }\mu\text{m}$, $450\text{ }\mu\text{m}$, and radio wavelengths. We note also that 1–34.0 shows no obvious signs of starburst activity at optical wavelengths, and appears to be a relatively undisturbed spiral galaxy. Should the identification with 1–34.2 prove to be erroneous, we note for completeness that two $z_{\text{ph}} \approx 3.9$ galaxies, the nearer being 1–27.0 and the further being 1–31.0, lie within 4 arcsec of HDF850.3. Based on magnitude, the probability that these are chance associations is 0.63, while based on redshift, it is 0.3.

HDF850.4 lies less than 1 arcsec from 2–339.0, an $I_{814}(AB) = 23$ galaxy for which photometric redshifts have been determined in the range 0.74–0.88 (refs 43, 47, 48). This is a convincing identification because the probability (based on its magnitude) that 2–339.0 is a chance association with HDF850.4 is only 0.07 (based on its redshift, the probability is 0.44). Moreover, this is the one case for

which the $850\text{-}\mu\text{m}$ source can be plausibly associated with an ISOCAM detection at $15\text{ }\mu\text{m}$, which yields a flux density ratio $S(850\text{ }\mu\text{m}/15\text{ }\mu\text{m}) \approx 16$. This is in fact exactly the value expected from the observed SED of Arp220 if placed at $z \approx 1$. This supports an identification with 2–339.0, and emphasizes the usefulness of constraining the redshift using the $S(850\text{ }\mu\text{m}/15\text{ }\mu\text{m})$ ratio. Furthermore, we note that this optical galaxy is clearly disturbed, as would be expected for an extreme starburst galaxy, providing further circumstantial evidence that it is the correct optical identification. However, should this identification prove erroneous, we note that there are three galaxies (2–294.0, 2–315.0, 2–319.0) with $z_{\text{ph}} > 3$ within 3.5 arcsec of HDF850.4.

HDF850.5 is located in a sparsely populated region of the HDF, but is only 0.9 arcsec away from the $I_{814}(AB) \approx 29$ galaxy 2–426.0, for which we estimate a photometric redshift of $z_{\text{ph}} = 3.2$. This is a moderately convincing identification; based on its magnitude, the probability that it is a chance coincidence is 0.46, but based on its redshift, it is a more impressive 0.16. This high-redshift association is consistent with the lack of a $15\text{-}\mu\text{m}$ detection at that position, but we note the difficulty of estimating photometric redshifts for such faint HDF galaxies. Finally, we note that seven other faint ($I_{814}(AB) > 27.5$) galaxies lie within 3.5 arcsec of HDF850.5 (see Fig. 4), but these objects are significantly more likely to be chance coincidences, and in any case all also have photometric redshifts $z_{\text{ph}} > 2$.

In summary, HDF850.4 appears to be associated with a disturbed starburst galaxy at $z_{\text{ph}} \approx 1$, HDF850.3 with an asymmetric galaxy at $z_{\text{ph}} \approx 2$, and HDF850.1 and HDF850.5 have relatively unambiguous associations with HDF galaxies, both of which are at $z_{\text{ph}} \approx 3$. For HDF850.2 we find a possible identification within the HDF at $z \approx 4$, and an alternative (also faint) candidate in the HFF which can be reasonably expected to lie at $z > 2$. We note that we are clearly unable to rule out the possibility that the true optical counterparts of a few of these sources may be too faint for detection, even in the HDF.

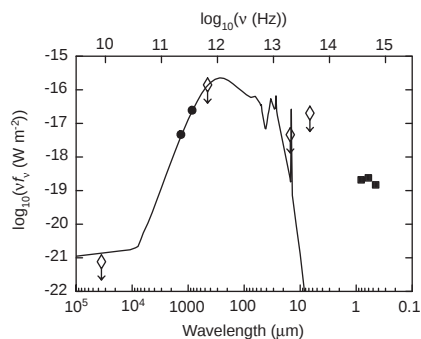


Figure 5 The observed optical-radio spectral energy distribution of HDF850.1. The filled circles represent the SCUBA 850- and 1,350- μm detections. Non-detections at 8.5 GHz (ref. 42), $450\text{ }\mu\text{m}$ (this work), 15 and $6.7\text{ }\mu\text{m}$ (S.O. *et al.*, manuscript in preparation) are shown as open diamonds. The filled squares indicate the optical fluxes in the I_{814} , V_{606} and B_{450} HST bands of the $z = 3.36$ galaxy 3–577.0, which is a plausible association for HDF850.1. A starburst galaxy model (solid curve)³⁸ has been redshifted to $z = 3.36$ and normalized to the $850\text{-}\mu\text{m}$ flux density. An additional radio non-thermal synchrotron component (where $F_{\nu} \propto \nu^{-0.8}$) is scaled to the starburst model at $60\text{ }\mu\text{m}$ using the radio/far-infrared correlation⁴¹.

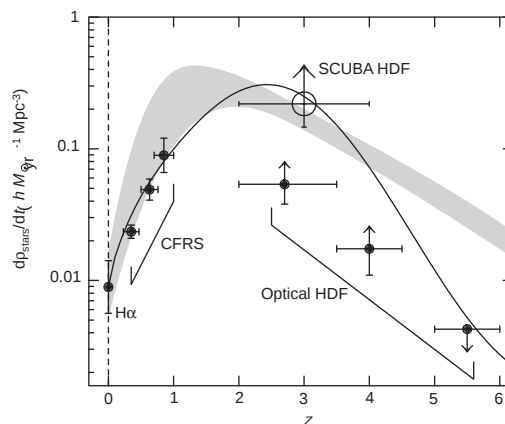


Figure 6 The global star-formation history of the Universe. Traditionally the mean co-moving rate of formation of stars in the universe, $d\rho_{\text{stars}}/dt$ has been measured from the total ultraviolet luminosity density of galaxies. At $z < 1$, this was measured by the Canada–France Redshift Survey of Lilly *et al.*,¹ and at higher redshifts from the optical HDF data². The zero-redshift datum was inferred from local emission-line galaxies⁵⁷. The shaded region shows the prediction (assuming $h = 0.65$) due to Pei and Fall⁵³ who argued using the observed column densities in quasar absorbers, plus the low metallicities in these systems, that the star-formation rate must have peaked between $z = 1$ and $z = 2$. The solid line illustrates what would happen if the star-formation rate tracked the total output of radio-loud AGN⁵¹. Based on the evidence which indicates that four of the five brightest submillimetre HDF sources lie at $2 < z < 4$, we infer a rate ~ 5 times higher than that obtained by Madau³, but in good agreement with the external predictions^{51,52} of the rate at these epochs.

Dust masses and star-formation rates

The photometric redshifts and suggested optical identifications for the submillimetre sources are consistent with the expectation that all galaxies detected in the 850- μm survey of the HDF down to a flux limit $S_{850\mu\text{m}} = 2\text{ mJy}$ should have redshifts $z \geq 1$. Given this, and the flat flux-density/redshift relation between $z = 1$ and $z = 10$, the dust-enshrouded SFRs and the dust masses for all five reliable sources can be estimated, independent of their precise redshifts. The results are given in Table 2, and indicate that these sources are extremely dusty and have SFRs, when determined from the submillimetre data, that are similar to, or exceed, that of the local ultra-luminous starburst galaxy Arp220. We note that the calculated SFRs are sensitive to the assumed stellar initial mass function (IMF) and stellar mass-range (and can, in the extreme increase and decrease by a factor of ~ 3). More striking is the comparison of the far-infrared SFRs with those calculated from the rest-frame ultraviolet luminosities. The far-infrared method gives SFRs on average a factor of ~ 300 larger. It has been shown that optical SFRs, estimated from Balmer emission-line luminosities in Lyman-break galaxies at $z \approx 3$, are larger than the ultraviolet SFRs by factors of 2–15. This upward correction, due to attenuation by dust^{9,10}, would still require a further factor of ~ 50 to explain the higher far-infrared SFRs. A similar situation has been observed in the local Universe where the ratio of far-infrared and H α luminosities in ultra-luminous infrared starburst galaxies (ULIRGs) is ~ 60 times larger than that in disk galaxies⁵⁰, suggesting that in young starbursts most OB stars are still deeply buried in their opaque parent clouds. The submillimetre sources we have seen are then quite typical of local ULIRGs, but their relevance to the overall cosmic star-formation history depends on their space density.

The small uncorrected ultraviolet SFRs ($< 1h^{-2}M_{\odot}\text{yr}^{-1}$, Table 2) for the optical counterparts of the submillimetre sources are reasonable, given that these galaxies are typically a few magnitudes fainter in the rest-frame ultraviolet, possibly due to greater dust obscuration, than the population of Lyman-break galaxies at $z \approx 3.5$ which have SFRs of $\sim 2h^{-2}M_{\odot}\text{yr}^{-1}$. Alternatively, it may be that the less-secure identifications are erroneous, and that some of the submillimetre sources may have true optical counterparts below the detection limit on the HST HDF image, and therefore probably at $z \geq 5$.

By summing the far-infrared SFRs and dividing by the appropriate cosmological volume, a first, conservative estimate of the level of dust-enshrouded star-formation rate in the high-redshift Universe can be made, using observations that are insensitive to the obscuring effects of dust. For illustrative purposes, it can be reasonably assumed that four of the five sources (all but HDF850.4) lie in the redshift interval $2 < z < 4$, in which case a lower-limit to the dust-enshrouded SFR density is $0.21hM_{\odot}\text{yr}^{-1}\text{Mpc}^{-3}$ (assuming an Einstein–de Sitter Universe) at $z \approx 3$. This datum is plotted in Fig. 6, where it can be compared with the optically derived star-formation history of the Universe^{3,4}, the dust-corrected star-formation history

predicted from the evolution of radio-loud active galactic nuclei⁵¹ and that inferred from the metal-production rate as determined from the observed column densities and metallicities in absorbers^{52,53} of radiation from quasars.

If the redshift distribution extended beyond $z = 4$, the mean redshift would increase, but due to increased cosmological volume the SFR density would decrease in such a way that the data would remain consistent with the curves shown in Fig. 6. For example, assuming a redshift range $2 < z < 5$ yields a star-formation density of $0.16hM_{\odot}\text{yr}^{-1}\text{Mpc}^{-3}$ at $z \approx 3.5$.

We emphasize that the SCUBA datum plotted in Fig. 6 is in fact rather robust. For example, should our proposed identification for HDF850.3 (1–34.2) prove to have a spectroscopic redshift significantly lower than $z = 2$, the effect on the data in Fig. 6 will be to lower the $z \approx 3$ SCUBA datum by only 20%. However, the effect of even one of these sources lying at still higher redshift is rather dramatic; if one of the 2–3 mJy submillimetre sources we have detected in fact lies at $z > 4$, this would yield a star-formation density of $0.1hM_{\odot}\text{yr}^{-1}\text{Mpc}^{-3}$ in the redshift range $4 < z < 6$, thus keeping the star-formation density essentially constant out to $z \approx 5$.

Finally, it is interesting to consider how the high-redshift star-formation density inferred from our submillimetre detections is altered by adoption of a much lower dust temperature (and hence much lower total far-infrared luminosity) than is displayed by ultra-luminous infrared starburst galaxies in the local Universe. The values of SFR given in column 5 of Table 2 have been derived using a starburst model of M82, and essentially identical numbers result from the adoption of a single dust temperature of 50 K. Nevertheless, it is certainly possible to reduce these values by up to a factor of ~ 10 if a single dust temperature of 20 K is adopted, and this would be sufficient to lower the $z = 3$ SCUBA datum plotted in Fig. 6 to a level consistent with the uncorrected star-formation density inferred from optical observations. However, such a situation must be regarded as extremely improbable for two reasons. First, it is highly unlikely that the most luminous far-infrared sources in the high-redshift Universe should have dust temperatures more comparable to that of the Milky Way than to all known luminous far-infrared galaxies. Second, while the adoption of such low dust temperatures reduces inferred SFR, it also inevitably results in much larger derived dust masses, > 10 times larger than the values listed in Table 2. The inferred molecular gas masses in the $z \approx 3$ galaxies are then so large ($\sim 5 \times 10^{11}h^{-2}M_{\odot}$) that a star-formation density of $> 0.5hM_{\odot}\text{yr}^{-1}\text{Mpc}^{-3}$ at higher redshift is then required to produce the necessary quantities of chemically enriched material⁵⁸; reference to Fig. 6 shows that such a value would be ~ 100 times greater than the star-formation density at $z > 4$ inferred from optical observations. While such a discovery would be extremely exciting, this alternative interpretation of our data would seem very unlikely to be correct. In summary, adoption of an average dust temperature of 50 K for the brightest submillimetre sources in our survey is most consistent with all the available evidence, and does not require inferred star-formation densities at $z > 4$ which differ by orders of magnitude from the external predictions illustrated in Fig. 6.

The deep submillimetre survey of the HDF reported here demonstrates that a significant fraction ($> 80\%$) of the star-formation activity in the high-redshift Universe may have been missed in previous optical studies. Four of the five brightest submillimetre sources alone provide a density of dust-enshrouded star-formation at $z > 2$ which is at least a factor of ~ 5 greater than that deduced from Lyman-limit systems³. The extent to which even this is an under-estimate depends on the number of sources fainter than $S_{850} = 2\text{ mJy}$ at comparable redshift.

Our survey has identified a population of high-redshift dusty starburst galaxies which contribute a significant fraction of the extragalactic background at 850 μm . These observations, together with complementary wider, shallower submillimetre surveys, are

Table 2 Star-formation rates and dust masses of the sources in Table 1

source	z_{est}	$\log_{10}L_{60\mu\text{m}}$ ($h^{-2}L_{\odot}$)	SFR ($h^{-2}M_{\odot}\text{yr}^{-1}$)		$\log_{10}M_{\text{dust}}$ ($h^{-2}M_{\odot}$)
			UV	FIR	
HDF850.1	3.4	12.15	0.7	311	7.88
HDF850.2	3.8	11.87	0.2	161	7.62
HDF850.3	2.0	11.76	2.0	127	7.52
HDF850.4	0.9	11.83	0.7	142	7.50
HDF850.5	3.2	11.64	0.3	95	7.36

Column 2, photometric redshifts based on the most probable optical associations. Column 3, 60- μm luminosity determined from the starburst model of M82³⁸ scaled to the observed 850- μm flux densities. Columns 4, 5, star-formation rates calculated from rest-frame ultraviolet (UV; 2,800 Å) and far infrared (FIR; 60 μm) luminosities^{52,53}. The UV flux densities at 2,800 Å are interpolated from the measured $I_{814}(AB)$ and $V_{606}(AB)$ magnitudes³⁹. Column 6, dust masses, assuming $\beta = 1.5$, $T = 50\text{ K}$. An Einstein–de Sitter cosmology is assumed.

now beginning to provide the first true measurement of the star-formation history for the early Universe, unhindered by the attenuating effects of dust. The challenge for the future is to follow up these observations, in particular those of the submillijansky sources which at present can only be detected statistically.

It is very possible that some of these submillimetre sources lie at $z > 5$, but demonstrating this will require the individual detection of these sources with sub-arcsec position errors. This will require both improved noise performance and higher spatial resolution in order to evade the confusion limit. Facilities such as the forthcoming generation of submillimetre arrays are ideally matched to these important programmes for astrophysical cosmology. □

Methods

The data, taken in jiggle-map mode, were reduced in parallel using two wholly independent methods. The first reduction used the SCUBA User Reduction Facility (SURF v1.2)⁵⁵, whilst the second reduction was performed with a specially-written IDL pipeline. Both methods incorporate individual bolometer r.m.s. noise weighting in the map reconstruction. An iterative temporal de-glitching and spatial sky subtraction was performed. The IDL maps were reconstructed using a noise-weighted “drizzling” technique⁵⁶ and were in excellent agreement with the independent SURF reconstructions. Individual subsets of the data were also reduced using both techniques, including the production of separate images with 30-arcsec and 45-arcsec chop throws. The centre of the map contains a higher density of bolometer samples than the periphery. Consequently, the noise is a function of position and this variation can be deduced exactly from the known jiggle pattern, and is approximated closely by a quadratic radial variation $\sigma \propto 1 + (r/90 \text{ arcsec})^2$ in the central regions. In the SURF reduction, the noise has the statistical character of white noise filtered with a beam of FWHM 6 arcsec, with an r.m.s. of 0.65 mJy at the map centre. Convolution of the map reduces this noise, but possible confusion from faint sources means that it is preferable not to broaden the point-source response significantly. As a compromise, a further convolution with a 6-arcsec beam was applied, reducing the r.m.s. noise signal to 0.45 mJy at the map centre. The noise on the final map therefore has the character of white noise convolved with a beam of FWHM 8.5 arcsec. A signal-to-noise image is shown in Fig. 1, allowing the significance of faint potential sources to be judged in a uniform manner, although it means that there is a tendency for more sources to be detected in the central regions. The analysis of sources was restricted to the central 80-arcsec radius. Because the observing strategy yields a point-source response with negative sidelobes at 0.25 of the peak, the map was CLEANed and restored with a 14.7-arcsec FWHM gaussian beam.

Received 26 May; accepted 2 July 1998.

1. Lilly, S. J., Le Fèvre, O., Hammer, F. & Crampton, D. The Canada-France redshift survey: the luminosity density and star formation history of the universe to $z \sim 1$. *Astrophys. J.* **460**, L1–L4 (1996).
2. Steidel, C. C. & Hamilton, D. Deep imaging of high redshift QSO fields below the Lyman limit. I. The field of Q0000–263 and galaxies at $z = 3.4$. *Astron. J.* **104**, 941–949 (1992).
3. Madau, P. et al. High-redshift galaxies in the Hubble Deep Field: colour selection and star formation history to $z \sim 4$. *Mon. Not. R. Astron. Soc.* **283**, 1388–1404 (1996).
4. Madau, P. The evolution of luminous matter in the universe. In *The Hubble Deep Field* (eds Livio, M. et al.) (STScI Symp. Ser., in the press).
5. Connolly, A. J., Szalay, A. S., Dickinson, M., SubbaRao, M. V. & Brunner, R. J. The evolution of the global starformation history as measured from the Hubble Deep Field. *Astrophys. J.* **486**, L11–L14 (1997).
6. Fall, S. M., Charlot, S. & Pei, Y. C. Cosmic emissivity and background intensity from damped Lyman-alpha galaxies. *Astrophys. J.* **464**, L43–L46 (1996).
7. Franceschini, A., Mazzei, P., De Zotti, G. & Danese, L. Luminosity evolution and dust effects in distant galaxies: implications for the observability of the early evolutionary phases. *Astrophys. J.* **427**, 140–154 (1994).
8. Pettini, M., Smith, L. J., King, D. L. & Hunstead, R. W. The metallicity of high-redshift galaxies: the abundance of zinc in 34 damped Ly- α systems from $z = 0.7$ to 3.4 . *Astrophys. J.* **486**, 665–680 (1997).
9. Pettini, M. et al. The spectra of star forming galaxies at high redshift. In *The Ultraviolet Universe at Low and High Redshift* (ed. Waller, W.) (AIP Press, Woodbury, New York, in the press).
10. Heckman, T. M., Robert, C., Leitherer, C., Garnett, D. R. & van der Rydt, F. The ultraviolet spectroscopic properties of local starbursts: implications at high-redshift. *Astrophys. J.* (submitted); also as preprint astro-ph/9803185 at (<http://xxx.lanl.gov>) (1998).
11. Meurer, G. R., Heckman, T. M., Lehnert, M. D., Leitherer, C. & Lowenthal, J. The panchromatic starburst intensity limit at low and high redshift. *Astron. J.* **114**, 54–68 (1997).
12. Franceschini, A., De Zotti, G., Toffolatti, L., Mazzei, P. & Danese, L. Galaxy counts and contributions to the background radiation from 1 micron to 1000 microns. *Astron. Astrophys. Suppl. Ser.* **89**, 285–310 (1991).
13. Blain, A. W. & Longair, M. S. Submillimetre cosmology. *Mon. Not. R. Astron. Soc.* **264**, 509–521 (1993).
14. Hughes, D. H., Dunlop, J. S. & Rawlings, S. High-redshift radio galaxies and quasars at sub-millimetre wavelengths: assessing their evolutionary status. *Mon. Not. R. Astron. Soc.* **289**, 766–782 (1997).

15. Rowan-Robinson, M. et al. The ultraviolet-to-radio continuum of the ultraluminous galaxy IRAS F10214+4725. *Mon. Not. R. Astron. Soc.* **261**, 513–521 (1993).
16. Barvainis, R., Antonucci, R., Hurt, T., Coleman, P. & Reuter, H.-P. The broadband spectral energy distributions of the Cloverleaf quasar and IRAS F10214+4724. *Astrophys. J.* **451**, L9–L12 (1995).
17. Dunlop, J. S., Hughes, D. H., Rawlings, S., Eales, S. A. & Ward, M. J. Detection of a large mass of dust in a radio galaxy at redshift $z = 3.8$. *Nature* **370**, 347–349 (1994).
18. Ivison, R. J. et al. Dust, gas, and evolutionary status of the radio galaxy 8C 1435+635 at $z = 4.25$. *Astrophys. J.* **494**, 211–217 (1998).
19. Isaak, K. G., McMahon, R. G., Hills, R. E. & Withington, S. Observations of high redshift objects at submillimetre wavelengths. *Mon. Not. R. Astron. Soc.* **269**, L28–L31 (1994).
20. Omont, A. et al. Continuum millimetre observations of high-redshift radio-quiet QSOs-II. Five new detections at $z > 4$. *Astron. Astrophys.* **315**, 1–10 (1996).
21. Holland, W. S. et al. in *Advanced Technology MMW, Radio, and TeraHertz Telescopes Proc. SPIE* **3357** (in the press).
22. Smail, I., Ivison, R. J. & Blain, A. W. A deep sub-millimetre survey of lensing clusters: a new window on galaxy formation and evolution. *Astrophys. J.* **490**, L5–L8 (1997).
23. Saunders, W. et al. The 60-micron and far-infrared luminosity functions of IRAS galaxies. *Mon. Not. R. Astron. Soc.* **242**, 318–337 (1990).
24. Lawrence, A. et al. The QDOT all-sky IRAS galaxy redshift survey. *Mon. Not. R. Astron. Soc.* (submitted).
25. Oliver, S. J. et al. Observations of the Hubble Deep Field with the Infrared Space Observatory-III. Source counts and P(D) analysis. *Mon. Not. R. Astron. Soc.* **289**, 471–481 (1997).
26. Rowan-Robinson, M. et al. Observations of the Hubble Deep Field with the Infrared Space Observatory-V. Spectral energy distributions, starburst models and star formation history. *Mon. Not. R. Astron. Soc.* **289**, 490–496 (1997).
27. Blain, A. W. & Longair, M. S. Observing strategies for blank-field surveys in the submillimetre waveband. *Mon. Not. R. Astron. Soc.* **279**, 847–858 (1996).
28. Pearson, C. & Rowan-Robinson, M. Starburst galaxy contributions to extragalactic source counts. *Mon. Not. R. Astron. Soc.* **283**, 174–192 (1996).
29. Hughes, D. H. & Dunlop, J. S. Continuum observations of the high-redshift universe at sub-millimetre wavelengths. In *Highly Redshifted Radio Lines* (eds Carilli, C. et al.) (ASP Conf. Ser., ASP, San Francisco, in the press).
30. Williams, R. E. et al. The Hubble Deep Field: observations, data reduction, and galaxy photometry. *Astron. J.* **112**, 1335–1389 (1996).
31. Bond, J. R. & Efstathiou, G. The statistics of cosmic background radiation fluctuations. *Mon. Not. R. Astron. Soc.* **226**, 655–687 (1987).
32. Condon, J. J. Confusion and flux-density error distributions. *Astrophys. J.* **188**, 279–286 (1974).
33. Steidel, C. C. et al. A large structure of galaxies at redshift $z \sim 3$ and its cosmological implications. *Astrophys. J.* (in the press); also as preprint astro-ph/9708125 at (<http://xxx.lanl.gov>) (1997).
34. Gialalisco, M. et al. The angular clustering of Lyman-break galaxies at redshift $z = 3$. *Astrophys. J.* (in the press); also as preprint astro-ph/9802318 at (<http://xxx.lanl.gov>) (1998).
35. Puget, J. L. et al. Tentative detection of a cosmic far-infrared background with COBE. *Astron. Astrophys.* **308**, L5–L8 (1996).
36. Fixen, D. J., Dwek, E., Mather, J. C., Bennett, C. L. & Shafer, R. A. The spectrum of the extra-galactic FIR background from the COBE IRAS observations. Preprint astro-ph/9803021 at (<http://xxx.lanl.gov>) (1998).
37. Guiderdoni, B., Bouchet, F. R., Puget, J. L., Lagache, G. & Hivon, H. The optically dark side of galaxy formation. *Nature* **390**, 257–259 (1998).
38. Rowan-Robinson, M. & Efstathiou, A. Multigrain dust cloud models of starburst and Seyfert galaxies. *Mon. Not. R. Astron. Soc.* **263**, 567–680 (1993).
39. Dudley, C. C. & Wynn-Williams, C. G. The deep silicate absorption feature in IRAS 08572+3915 and other infrared galaxies. *Astrophys. J.* **488**, 720–729 (1997).
40. Ivison, R. J. et al. A hyperluminous galaxy at $z = 2.8$ found in a deep submillimetre survey. *Mon. Not. R. Astron. Soc.* (in the press); also as preprint astro-ph/9712161 at (<http://xxx.lanl.gov>) (1997).
41. Sopp, H. M. & Alexander, P. A composite plot of far-infrared versus radio luminosity, and the origin of far-infrared luminosity in quasars. *Mon. Not. R. Astron. Soc.* **251**, 14P–16P (1991).
42. Richards, E. A., Kellermann, K. I., Fomalont, E. B., Windhorst, R. A. & Partridge, R. B. Radio emission from galaxies in the Hubble Deep Field. *Astron. J.* (submitted); also as preprint astro-ph/9803343 at (<http://xxx.lanl.gov>) (1998).
43. Mobasher, B., Rowan-Robinson, M., Georgakakis, A. & Eaton, N. The nature of the faint galaxies in the Hubble Deep Field. *Mon. Not. R. Astron. Soc.* **282**, L7–L14 (1996).
44. Browne, I. W. A. & Cohen, A. M. Quasars near bright galaxies. *Mon. Not. R. Astron. Soc.* **182**, 181–187 (1978).
45. Zepf, S. E., Moustakas, L. A. & Davis, M. Keck spectroscopy of objects with lens-like morphologies in the Hubble Deep Field. *Astrophys. J.* **474**, L1–L5 (1997).
46. Hogg, D. W., Blandford, R., Kundic, T., Fassnacht, C. D. & Malhotra, S. A candidate gravitational lens in the Hubble Deep Field. *Astrophys. J.* **467**, L73–L75 (1996).
47. Sawicki, M. J., Lin, H. & Yee, H. K. C. Evolution of the galaxy population based on photometric redshifts in the Hubble Deep Field. *Astron. J.* **113**, 1–12 (1997).
48. Wang, Y., Bahcall, N. & Turner, E. L. A catalogue of colour-based redshift estimates for $z \leq 4$ galaxies in the Hubble Deep Field. *Astron. J.* (submitted); also as preprint astro-ph/9804195 at (<http://xxx.lanl.gov>) (1998).
49. Oke, J. B. Absolute spectral energy distributions for white dwarfs. *Astrophys. J. Suppl.* **27**, 21–35 (1974).
50. Leech, K. J. High-luminosity IRAS galaxies-II. Optical spectroscopy, modelling of starburst regions and comparison with structure. *Mon. Not. R. Astron. Soc.* **240**, 349–372 (1989).
51. Dunlop, J. in *Observational Cosmology with the New Radio Surveys* (eds Bremer, M. N. et al.) 157–164 (Kluwer, Dordrecht, 1998).
52. Fall, S. M. & Pei, Y. C. Obscuration of quasars by dust in damped Lyman-alpha systems. *Astrophys. J.* **402**, 479–492 (1993).
53. Pei, Y. C. & Fall, S. M. Cosmic chemical evolution. *Astrophys. J.* **454**, 69–76 (1995).
54. Sandell, G. Secondary calibrators at submillimetre wavelengths. *Mon. Not. R. Astron. Soc.* **271**, 75–80 (1994).
55. Jenness, T. & Lightfoot, J. F. in *Astronomical Data Analysis Software and Systems VII* (eds Albrecht, R. et al.) 216–219 (ASP Conf. Ser. 145, ASP, San Francisco, 1998).
56. Hook, R. N. & Fruchter, A. S. in *Astronomical Data Analysis Software and Systems VI* (eds Hunt, G. & Payne, H. E.) 147 (ASP Conf. Ser. 125, ASP, San Francisco, 1997).
57. Gallego, J., Zamorano, J., Aragon-Salamanca, A., Rego, M. The current starformation rate of the local universe. *Astrophys. J.* **455**, L1–L4 (1995).
58. Eales, S. A. & Edmunds, M. G. The implications of large dust masses at high redshifts: a first look at galactic evolution in the submillimetre waveband. *Mon. Not. R. Astron. Soc.* **280**, 1167–1180 (1996).

Correspondence and requests for materials should be addressed to D.H. (e-mail: D.Hughes@roe.ac.uk).

Submillimetre-wavelength detection of dusty star-forming galaxies at high redshift

A. J. Barger*, L. L. Cowie*, D. B. Sanders*, E. Fulton*, Y. Taniguchi†, Y. Sato‡§, K. Kawarall & H. Okuda§

* Institute for Astronomy, University of Hawaii, 2680 Woodlawn Drive, Honolulu, Hawaii 96822, USA

† Astronomical Institute, Tohoku University, Aoba, Sendai 980-8578, Japan

‡ ISO Science Operations Centre, Astrophysics Division of ESA, Villafranca, E-28080, Madrid, Spain

§ Institute of Space and Astronautical Science, Yoshinodai, Sagamihara, Kanagawa 229, Japan

|| Institute of Astronomy, University of Tokyo, 2-21-1 Osawa, Mitaka, Tokyo 181, Japan

Optical surveys of the global star-formation rate in high-redshift galaxies show a strong peak in activity at a redshift of $z \approx 1.5$, which implies that most of the star formation¹ has already been seen. High-redshift galaxies may, however, emit most of their energy at submillimetre wavelengths, if they contain substantial amounts of dust that absorbs the starlight and reradiates it as far-infrared light. Here we report a deep survey of a blank region of sky, performed at submillimetre wavelengths (450 and 850 μm). We detect luminous sources in the 850- μm band which, if they have similar spectra to low-redshift ultraluminous infrared galaxies and are primarily powered by star formation, must each be converting more than 100 solar masses of gas per year into stars: this is larger than the maximum star-formation rates inferred for most optically selected galaxies². The total amount of star formation at high redshifts is essentially fixed by the level of background light, but where the peak activity occurs at submillimetre wavelengths is not yet well established. However, the background light inferred from the sources that we have detected is already comparable to that from the optically selected sources. Establishing the main epoch of star formation will therefore require a combination of optical and submillimetre studies.

In recent years high-redshift optical galaxy searches have become increasingly successful at uncovering significant populations of galaxies that are likely to be in early phases of evolution. However, the global star formation rate (SFR) inferred^{1,3,4} omits the many fainter sources that are now being detected^{5,6}. Furthermore, the effects of dust can cause the SFRs in the detected ultraviolet-bright objects to be grossly underestimated (see, for example, ref. 7), and many rapidly star forming galaxies may even be omitted from the optical samples.

Nearby star-forming galaxies emit a large fraction of their bolometric luminosity in the far-infrared waveband, which for distant sources is redshifted into the submillimetre waveband. Because the far-infrared spectra of these star-forming galaxies are very steep, if they are at large redshifts their flux density decreases much less rapidly with increasing redshift than is expected according to the inverse-square law. Thus, galaxies at high redshifts are likely to make a substantial contribution to the submillimetre counts⁸. In a pioneering paper⁹, Smail *et al.* reported the detections of submillimetre sources in the fields of two rich clusters of galaxies using the new SCUBA bolometer array¹⁰ on the 15-m James Clerk Maxwell Telescope. These clusters were selected because they are strong gravitational lenses which magnify any submillimetre sources lying behind them. The objects in the A370 field include one spectacular source, SMM02399-136, with an 850- μm flux of 26 mJy ($1 \text{ Jy} = 1.0 \times 10^{-23} \text{ erg s}^{-1} \text{ cm}^{-2} \text{ Hz}^{-1}$). This source has subsequently been shown¹¹ to be an active galactic nucleus

embedded in a starburst at $z = 2.8$. Gravitational lensing greatly improves source detectability but entails an uncertain correction to remove the estimated magnification of the sources. The most likely amplification factor for the above source was given¹¹ as 2.5 with a firm upper limit of 5. The magnification problem can be avoided if blank field observations are made with longer exposure times. That is the nature of the experiment reported here.

We obtained SCUBA observations of two blank field areas, one in the Lockman hole and one in the Hawaii deep-field region SSA13. Our data were taken during eight 8-h shifts in March 1998 and eight 8-h shifts in May 1998. Fully sampled wavelength maps were completed simultaneously in both the long (850- μm) and short (450- μm) wavelength arrays, but the sky at 450 μm is considerably more opaque. Although two of the 16 shifts were too poor to

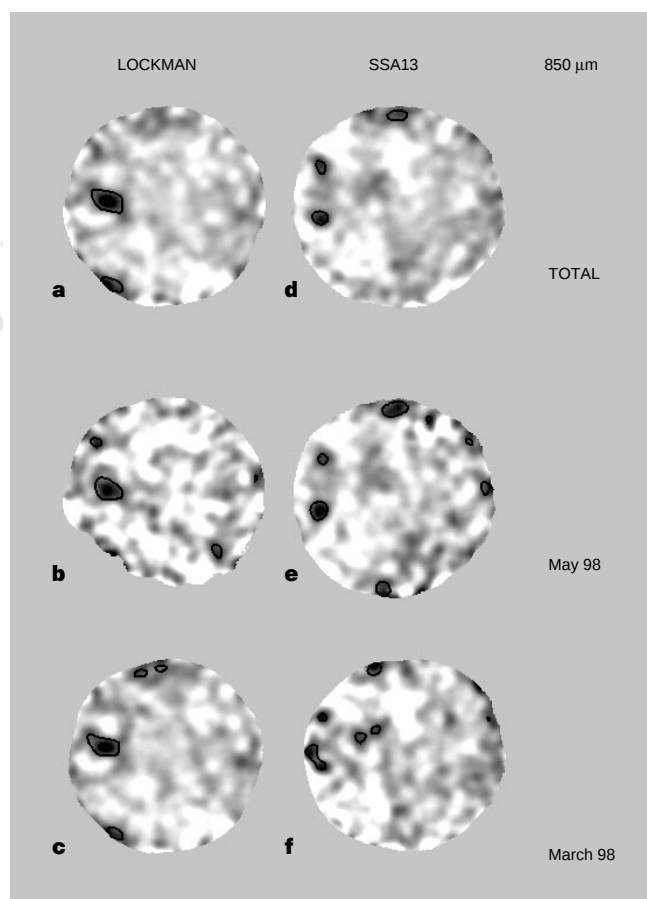


Figure 1 850- μm SCUBA maps of our two blank fields; Lockman hole (left) and SSA13 (right). These fields were chosen in areas of deep 7 μm surveys²⁹. The top panels show the total integrations, the middle panels show just the data from our May 1998 run, and the bottom panels show just the data from our March 1998 run. The maps have their centres aligned in the vertical. The field centres for the Lockman hole are right ascension (RA) 10 h 33 min 55.45 s and declination (dec.) $57^{\circ}46'18''$ (J2000) and for SSA13 are RA 13 h 12 min 25.6 s and dec. $42^{\circ}44'38''$ (J2000). The diameter of each of the maps is ~ 2.7 arcmin. The exposure times for the maps (in hours) are 38.8 (a), 8.1 (b), 30.7 (c), 51.0 (d), 34.8 (e), and 16.2 (f). The contours show levels of constant surface brightness. In the central regions of the Lockman hole maps, the first contour level is 2.2 mJy or $\sim 3\sigma$, and the second contour level is 4.4 mJy or $\sim 6\sigma$. In the central regions of the SSA13 maps, the contour level is 2.2 mJy or $\sim 3\sigma$. Any features right at the edges of the maps are insignificant at even the 1.5σ level due to the rapid increase in the noise levels at the very edge. Therefore, the only feature in all the maps (other than the two sources listed in Table 1) that we feel might be real is the 2.5-mJy feature ~ 45 arcsec north of the identified source in the SSA13 map. See Methods for details of data acquisition.

contribute significantly to the observations, much of the run had superb weather conditions and produced sensitive maps at both wavelengths. For the 850- μm maps, which are of primary interest, the total integration time was 39 hours on the Lockman hole region and 51 hours on SSA13.

At 450 μm , no sources were detected at the 3σ level of 25 mJy in either field. At 850 μm , two sources having fluxes >3 mJy were detected at the $>3\sigma$ level, one in each field; these are listed in Table 1. The 850- μm maps are described and shown in Fig. 1. The absolute value of the noise was found from the dispersion of the points in the map, excluding the regions around the brightest sources (see Table 1). Based on predictions of source confusion¹², the instrumental and sky noise are expected to dominate the total noise. The completeness of source recovery as a function of flux was tested using Monte Carlo simulations. Five sources at each of a range of flux levels were simulated using the beam derived from the calibration source and were added into the field. The source was considered to be recovered if it was detected at the 3σ level by the source detection algorithm. This procedure was repeated a large number of times at each flux. Recovery of sources brighter than 3 mJy was essentially complete, except for a small number of cases where the objects overlapped. The fall-off in recovery was extremely rapid, with 80% of the sources recovered around 3 mJy and only 20% by 2 mJy. Inspection of the negative images yielded no sources with fluxes greater than 3 mJy. Given that we detect only two sources in our data maps with fluxes above 3 mJy in 9 arcmin² of field, the cumulative surface density of sources above this flux is 800 (290–1,900) deg⁻², where the numbers in parentheses are the $\pm 1\sigma$ range. Our number is smaller than the tentative estimate of $2,500 \pm 1,400$ deg⁻² above 4 mJy made in ref. 9 on the basis of observations of

cluster-lensed sources. This difference may partially reflect the uncertainty in the magnification correction but could also arise in large part from the small-number statistics of the observations. Our value corresponds to a νS_ν , where S_ν is the sky surface brightness at frequency ν , of 3.7×10^{-8} erg cm⁻² s⁻¹ sr⁻¹, which is a factor of a few less than current limits on the extragalactic background light^{13–17}.

The strongest source in the two fields is the 4.6 mJy ($\sim 6\sigma$) source in the Lockman hole. In Fig. 2 we show a deep K' image of the Lockman hole SCUBA field with a 3 mJy per beam contour superimposed on the bright source. We also show deep zoomed K' and B -band images of the region surrounding the source. An optical spectrum obtained with the Low Resolution Imaging Spectrometer (LRIS¹⁸) on the Keck 10-m telescope reveals only continuum emission. Because [O II] 3,727 Å would shift out of the observed spectral range for $z \gtrsim 1.5$, this suggests that the source lies above this redshift limit. However, the presence of the objects in the B -band image places an upper limit on the redshift of the source of $z \lesssim 3.5$.

Table 1 850- μm sources

Field	RA (2000) (h mins)	Dec. (2000) (° ' ")	Flux (mJy)	1σ noise level at source position (mJy)
Lockman hole	10 34 02.3	57 46 25.0	4.6	0.8
SSA13	13 12 32.1	42 44 28.0	3.3	0.9

The UT 1998 dates and zenith 850- μm sky optical-depth ranges for the March run were 17 (0.30–0.31), 19 (0.42–0.47), 20 (0.20–0.35), 21 (0.13–0.14), 22 (0.13–0.30), 23 (0.13–0.24), 24 (0.10–0.11), 25 (0.10), and for the May run were 16 (0.28–0.41), 17 (0.27–0.36), 18 (0.28–0.38), 19 (0.45–0.58), 22 (0.11–0.14), 23 (0.11–0.14), 24 (0.20–0.23) and 25 (0.16–0.25). During the March run the airmass range for the Lockman hole was 1.3–2.4, and for SSA13 it was 1.1–1.4. For the May run, the airmass range for the Lockman hole was 1.3–1.5, and for SSA13 it was 1.1–1.7.

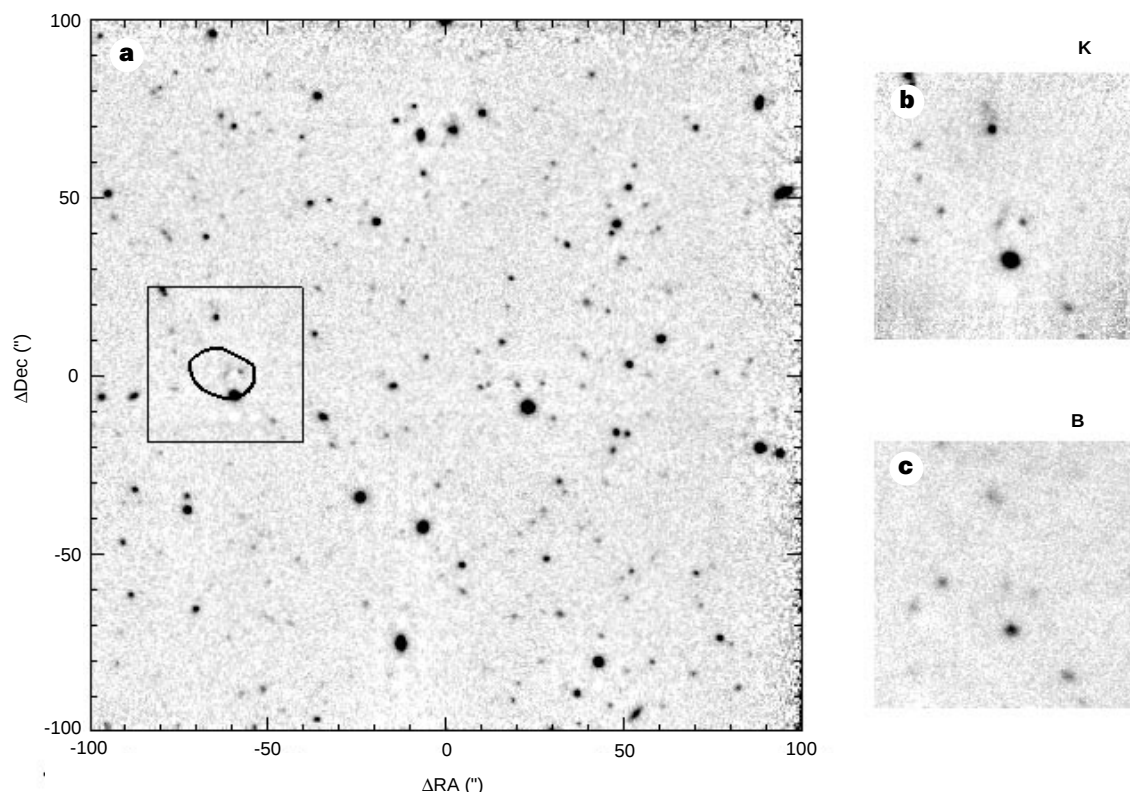


Figure 2 K' - and B -band images of the brightest source. **a**, 3 mJy per beam contour of the $\sim 6\sigma$ source detected in the 850- μm Lockman hole map displayed on a K' image obtained with the QUick Infrared Camera (QUIRC²⁷) on the University of Hawaii 88-inch telescope on Mauna Kea during the period of 1995–96. The units on the axes are arcseconds, and the square box is the 'zoomed' region shown in **b**. **b**, Zoomed K' image of the source (field centre)

obtained with the Near Infrared Camera (NIRC²⁸) on the Keck 10-m telescope in March 1998; $K'_{AB} = 21.8$. The displayed field is ~ 45 arcsec on a side. **c**, Zoomed B -band image obtained with the University of Hawaii 88-inch telescope in March 1998; $B_{AB} = 23.5$. The displayed field is ~ 45 arcsec on a side. Magnitudes were measured in 6-arcsec-diameter apertures and put into the AB magnitude system ($AB = -2.5 \log f_\nu - 48.60$, where f_ν is the flux in erg cm⁻² s⁻¹ Hz⁻¹).

Given these redshift constraints, the only class of local objects whose spectral energy distribution (SED) could provide an approximate match to the K' and 850 μm flux measurements of our brightest source is an ultraluminous infrared galaxy¹⁹. For the following discussion, we assume that the SED of the “prototypical” ultraluminous infrared galaxy Arp220 is appropriate for our object. The spectral shape of Arp220 can be represented²⁰ by a modified black body with emissivity λ^{-1} and a dust temperature of $T = 47$ K. If we place an Arp220 source at a redshift z , the effective black-body temperature becomes $T/(1+z)$ due to the expansion of the Universe. The far-infrared luminosity (L_{FIR}) can then be related to the observed 850- μm flux of our brightest source. Assuming the Arp220 temperature of $T = 47$ K, we find $L_{\text{FIR}} = (0.9\text{--}1.6) \times 10^{12} L_{\odot} h_{75}^{-2}$ for $q_0 = 0.5$ and $(1.2\text{--}6.1) \times 10^{12} L_{\odot} h_{75}^{-2}$ for $q_0 = 0.02$ over the redshift range $0.5 < z < 3.5$ (here $L_{\odot}$ is the solar luminosity, h_{75} is the Hubble constant in units of $75 \text{ km s}^{-1} \text{ Mpc}^{-1}$ and q_0 is the deceleration parameter). Submillimetre observations of nearly all luminous infrared galaxies have been reasonably fitted by single-temperature dust models¹⁹ with $T = 30\text{--}50$ K; thus, the temperature dependence of L_{FIR} introduces a factor of only $\sim 0.2\text{--}1.2$ uncertainty in the above L_{FIR} values.

Although we do not have radio spectral indices or diagnostic spectral-line ratios that are traditionally used to discriminate between thermal and non-thermal energy sources, it is probable—as in the low- z ultraluminous infrared galaxies—that the energy in our submillimetre sources is powered by a mixture of active galactic nuclei and star-formation, with a substantial fraction arising from the latter. If we assume that all of the energy originally in the rest-frame ultraviolet is re-radiated into the far-infrared and take the extreme case of zero contribution to the far-infrared emission from an obscured active galactic nucleus, then the far-infrared luminosity provides a measure of the current SFR of massive stars^{21,22} $\text{SFR} = \psi 10^{-10} (L_{\text{FIR}}/L_{\odot}) M_{\odot} \text{ yr}^{-1}$, where $\psi = 0.8\text{--}2.1$. To produce all of the luminous energy of our brightest source from intense continuing star formation would therefore require SFRs of $(140\text{--}250) M_{\odot} \text{ yr}^{-1} h_{75}^{-2}$ for $q_0 = 0.5$ and $(180\text{--}910) M_{\odot} \text{ yr}^{-1} h_{75}^{-2}$ for $q_0 = 0.02$, taking $\psi = 1.5$. This is an order of magnitude higher than the SFRs typically seen in rapidly star-forming galaxies observed in the optical^{2,3} and is similar to rates found in other recently targeted submillimetre observations (see, for example, ref. 23). Thus, our data indicate that the upper end of the mass formation function is dominated by objects that are radiating primarily in the submillimetre waveband.

The relative fraction of the universal stellar energy which is re-radiated into the submillimetre waveband and is present in these bright submillimetre sources, compared with that which is radiated in the rest-frame ultraviolet, is most simply quantified using the extragalactic background light. If we assume an Arp220 spectrum for the submillimetre objects and a substantial fraction of light powered by star formation, then the sum of our two measured $>3\sigma$ 850- μm source fluxes, if located at $z = 1$, would require re-radiation of a rest-frame ultraviolet sky brightness of $S_{\nu} = 9 \times 10^{-25} \text{ erg cm}^{-2} \text{ s}^{-1} \text{ Hz}^{-1} \text{ deg}^{-2}$ to produce the bolometric submillimetre light. (The redshift dependence of S_{ν} is approximately $((1+z)/2)^{-2.3}$). In comparison, the integrated flat-spectrum population corresponding to the peak of the optically selected star forming galaxies near $z = 1$ gives a rest-frame ultraviolet surface brightness²⁴ of approximately $4 \times 10^{-25} \text{ erg cm}^{-2} \text{ s}^{-1} \text{ Hz}^{-1} \text{ deg}^{-2}$. Although the submillimetre results are still quite uncertain due to the small number of sources, these two values are quite comparable, with the submillimetre radiation from the bright sources dominating if they lie at $z \lesssim 1.7$. More accurate estimates would require the observations of many SCUBA fields to the level described here together with spectroscopically determined redshifts and additional colour information.

Our results suggest that high-redshift star-forming galaxies with the largest SFRs radiate primarily in the submillimetre waveband.

Further support for this argument has been obtained in very recent cluster observations²⁵. Although these submillimetre sources with high SFRs are less common than are optically selected galaxies, the integrated radiation emitted by the submillimetre sources and by the optically selected galaxies seems to be comparable within the still-considerable uncertainties in the surface densities. The present results probe only the highest-luminosity submillimetre sources, and recent measurements of the integrated backgrounds indicate that several times more energy could be present in fainter submillimetre sources^{13–17}. □

Methods

SCUBA observations were obtained by stepping the secondary mirror through a 64-point jiggle pattern in Nasmyth coordinates while chopping at a frequency of 6.94 Hz and a chop throw of 45 arcsec in RA. In order to cancel out the sky, the jiggle pattern was subdivided and the target position switched between the signal and reference beams using a repeating signal-reference-reference-signal scheme. Because the chop throw is small the reference negative images also appear on the array and hence can be restored, increasing the effective exposure times on most of the field. The primary source extraction was done using beam- and exposure-weighting to determine the signal and noise at each spatial position, and both positive and negative portions of the beam pattern were used. Pointing checks were performed every hour using the blazars 0923+392 or 1308+326, and our data maps were calibrated using twice-nightly beam maps of the calibration source CRL618 (March run) or thrice-nightly beam maps of IRC+10216 (May run). The measurement of the calibration source was found to vary by a maximum of $\sim 20\%$ over the course of a shift. ‘Skydips’, which measure the zenith atmospheric opacities at 450 and 850 μm , were done every few hours throughout each night, and the sky opacity at 230 GHz was monitored at all times to keep track of the stability of the sky. The data were reduced in a standard way using the dedicated SCUBA data reduction software SURF²⁶.

Received 7 May; accepted 2 July 1998.

1. Madau, P. *et al.* High-redshift galaxies in the Hubble Deep Field: colour selection and star formation history to $z \sim 4$. *Mon. Not. R. Astron. Soc.* **283**, 1388–1404 (1996).
2. Pettini, M. *et al.* The discovery of primeval galaxies and the epoch of galaxy formation. In *Cosmic Origins: Evolution of Galaxies, Stars, Planets, and Life* (eds Shull, J. M., Woodward, E. E. & Thronson, H. A.) *Astron. Soc. Pacif.* (in the press); also as preprint astro-ph/9708117 at (<http://xxx.lanl.gov>) (1998).
3. Steidel, C. C., Giavalisco, M., Pettini, M., Dickinson, M. & Adelberger, K. L. Spectroscopic confirmation of a population of normal star-forming galaxies at redshifts $z > 3$. *Astrophys. J.* **462**, L17–L21 (1996).
4. Cowie, L. L., Hu, E. M., Songaila, A. & Egami, E. The evolution of the distribution of star formation rates in galaxies. *Astrophys. J.* **481**, L9–L14 (1997).
5. Cowie, L. L. & Hu, E. M. High- z Lyman-alpha emitters. I. a blank-field search for objects near redshift $z = 3.4$ in and around the Hubble Deep Field and the Hawaii Deep Field SSA22. *Astron. J.* **115**, 1319–1328 (1998).
6. Hu, E. M., Cowie, L. L. & McMahon, R. G. The density of Lyman-alpha emitters at very high redshift. *Astrophys. J.* (in the press); also as preprint astro-ph/9803011 at (<http://xxx.lanl.gov>) (1998).
7. Heckman, T. M., Robert, C., Leitherer, C., Garnett, D. R. & van der Rydt, F. The ultraviolet spectroscopic properties of local starbursts: implications at high-redshift. *Astrophys. J.* (in the press); also as preprint astro-ph/9803185 at (<http://xxx.lanl.gov>) (1998).
8. Blain, A. W. & Longair, M. S. Submillimeter cosmology. *Mon. Not. R. Astron. Soc.* **264**, 509–521 (1993).
9. Smail, I., Ivison, R. J. & Blain, A. W. A deep submillimeter survey of lensing clusters: a new window on galaxy formation and evolution. *Astrophys. J.* **490**, L5–L8 (1997).
10. Holland, W. S. *et al.* SCUBA—a submillimeter camera operating on the James Clerk Maxwell Telescope. In *Advanced technology MMW, radio and terahertz telescopes*. (ed. Phillips, T.) *Proc. SPIE* **3357**, (in the press).
11. Ivison, R. J. *et al.* A hyperluminous galaxy at $z = 2.8$ found in a deep submillimeter survey. *Mon. Not. R. Astron. Soc.* (in the press); also as preprint astro-ph/9712161 at (<http://xxx.lanl.gov>) (1998).
12. Blain, A. W., Ivison, R. J. & Smail, I. Observational limits to source confusion in the millimeter/submillimeter waveband. *Mon. Not. R. Astron. Soc.* **296**, L29–L33 (1998).
13. Puget, J. L. *et al.* Tentative detection of a cosmic far infrared background with COBE. *Astron. Astrophys.* **308**, L5–L8 (1996).
14. Burigana, C. *et al.* Constraints on the cosmic star formation history from the far-infrared background. *Mon. Not. R. Astron. Soc.* **287**, L17–L20 (1997).
15. Fixsen, D. J., Dwek, E., Mather, J. C., Bennett, C. L. & Shafer, R. A. The spectrum of the extragalactic far infrared background from the COBE FIRAS observations. *Astrophys. J.* (in the press); also as preprint astro-ph/9803021 at (<http://xxx.lanl.gov>) (1998).
16. Schlegel, D., Finkbeiner, D. P. & Davis, M. Maps of dust IR emission for use in estimation of reddening and CMBR foregrounds. *Astrophys. J.* **500**, 525–553 (1998).
17. Hauser, M. G. *et al.* The COBE diffuse infrared background experiment search for the cosmic infrared background: I. limits and detections. *Astrophys. J.* (in the press); also as preprint astro-ph/9806167 at (<http://xxx.lanl.gov>) (1998).
18. Oke, J. B., *et al.* The Keck low-resolution imaging spectrometer. *Publ. Astron. Soc. Pacif.* **107**, 375–385 (1995).
19. Sanders, D. B. & Mirabel, I. F. Luminous infrared galaxies. *Annu. Rev. Astron. Astrophys.* **34**, 749–782 (1996).

20. Klaas, U., Ilaas, M., Heinrichsen, I. & Schulz, B. Infrared spectral energy distributions of the interacting galaxies Arp 244, NGC 6240, and Arp 220. *Astron. Astrophys.* **325**, L21–L24 (1997).
21. Scoville, N. Z. & Young, J. S. The molecular gas distribution in M51. *Astrophys. J.* **265**, 148–165 (1983).
22. Thronson, H. & Telesco, C. Star formation in active dwarf galaxies. *Astrophys. J.* **311**, 98–112 (1986).
23. Cimatti, A., Andreani, P., Röttgering, H. & Tilanus, R. Vigorous star formation hidden by dust in a galaxy at $z = 1.4$. *Nature* **392**, 895–896 (1998).
24. Songaila, A., Cowie, L. L. & Lilly, S. J. Galaxy formation and the origin of the ionizing flux at large redshift. *Astrophys. J.* **348**, 371–377 (1990).
25. Blain, A. W., Smail, I., Ivison, R. J. & Kneib, J.-P. The history of star formation in dusty galaxies. *Mon. Not. R. Astron. Soc.* (submitted); also as preprint astro-ph 9806062 at (<http://xxx.lanl.gov>) (1998).
26. Jenness, T. *SURF—SCUBA User Reduction Facility* (Starlink User Note no. 216.1, J.A.C., Hilo, 1997).
27. Hodapp, K.-W. et al. The HAWAII infrared detector arrays: testing and astronomical characterization of prototype and science-grade devices. *New Astron.* **1**, 177–196 (1996).
28. Matthews, K. & Soifer, B. T. in *Infrared Astronomy with Arrays: the Next Generation* (ed. McLean, I.) 239–246 (Kluwer, Dordrecht, 1994).
29. Taniguchi, Y. et al. ISOCAM 7 micron deep survey of the Lockman hole: A mid-infrared search for primeval galaxies. *Astron. Astrophys.* **328**, L9–L12 (1997).

Acknowledgements. We thank R. Tilanus, T. Jenness, I. Smail and W. Vacca for discussions.

Correspondence and requests for materials should be addressed to A.J.B. (e-mail: barger@ifa.hawaii.edu).

A three-dimensional photonic crystal operating at infrared wavelengths

S. Y. Lin*, J. G. Fleming*, D. L. Hetherington*, B. K. Smith*, R. Biswas†, K. M. Ho†, M. M. Sigalas†, W. Zubrzycki*, S. R. Kurtz* & Jim Bur*

* Sandia National Laboratories, PO Box 5800, Albuquerque, New Mexico 87185, USA

† Ames Laboratory, Department of Physics and Astronomy, Iowa State University, Ames, Iowa 50011, USA

The ability to confine and control light in three dimensions would have important implications for quantum optics and quantum-optical devices: the modification of black-body radiation, the localization of light to a fraction of a cubic wavelength, and thus the realization of single-mode light-emitting diodes, are but a few examples^{1–3}. Photonic crystals—the optical analogues of electronic crystal—provide a means for achieving these goals. Combinations of metallic and dielectric materials can be used to obtain the required three-dimensional periodic variations in dielectric constant, but dissipation due to free carrier absorption will limit application of such structures at the technologically useful infrared wavelengths⁴. On the other hand, three-dimensional photonic crystals fabricated in low-loss gallium arsenide show only a weak ‘stop band’ (that is, range of frequencies at which propagation of light is forbidden) at the wavelengths of interest⁵. Here we report the construction of a three-dimensional infrared photonic crystal on a silicon wafer using relatively standard microelectronics fabrication technology. Our crystal shows a large stop band (10–14.5 μm), strong attenuation of light within this band (~ 12 dB per unit cell) and a spectral response uniform to better than 1 per cent over the area of the 6-inch wafer.

To create a three-dimensional (3D) photonic crystal, the so-called layer-by-layer periodic structure was adopted for its design simplicity^{6–10}. A diagram of the layer structure is shown in Fig. 1a. It consists of layers of one-dimensional rods with a stacking sequence that repeats itself every four layers with a repeat distance of c . Within each layer, the axes of the rods are parallel to each other with a pitch of d . The orientations of the axes are rotated by 90° between adjacent layers. Between every other layer, the rods are shifted relative to each other by $0.5d$. The resulting structure has a face-centred-tetragonal (f.c.t.) lattice symmetry. For the special case of $c/d = 1.414$, the lattice can be derived from a face-centred-cubic (f.c.c.) unit cell with a basis of two rods. This layered structure can also be derived by replacing the $\langle 110 \rangle$ chains of atoms in the diamond structure with the rods.

The layer structure shown in Fig. 1a shows a full 3D photonic bandgap, that is, a range of frequencies, or wavelengths, over which the photonic density of states (DOS) vanishes and the propagation of light is strictly forbidden along any direction. Such a gap is also called an ‘absolute gap’ as it blocks light from all directions. The gap size is determined by the dielectric contrast of the two different materials that constitute the 3D structure and by the filling fraction of the higher dielectric material. Our structure has a refractive index contrast of 3.60, defined as the ratio of the refractive index of Si ($n = 3.60$) and the refractive index of air ($n = 1.0$). Its filling fraction is given by $f = w/d = 0.28$, where w is the rod width. A computed photonic DOS plot summed over the entire Brillouin zone for such a 3D lattice is shown in Fig. 1b. The result shows a wide photonic bandgap extending from $0.46c_0/a$ to $0.56c_0/a$, where $a = 1.414d$ and c_0 is the speed of light *in vacuo*. The gap is larger in the stacking direction with band edges at $0.41c_0/a$ and $0.58c_0/a$.

The vertical topology of the 3D lattice structure is built by the repetitive deposition and etching of multiple dielectric films. To accomplish this goal, a comprehensive five-level stacking process was developed. Within each layer, SiO_2 was first deposited, patterned, and etched to the desired depth. The resulting trenches were then filled with polycrystalline silicon. Following this, the surface of the wafers were made flat using chemical mechanical polishing¹¹, and the process was then repeated. After the five-level process was completed, the wafer was immersed in a HF/water solution for the

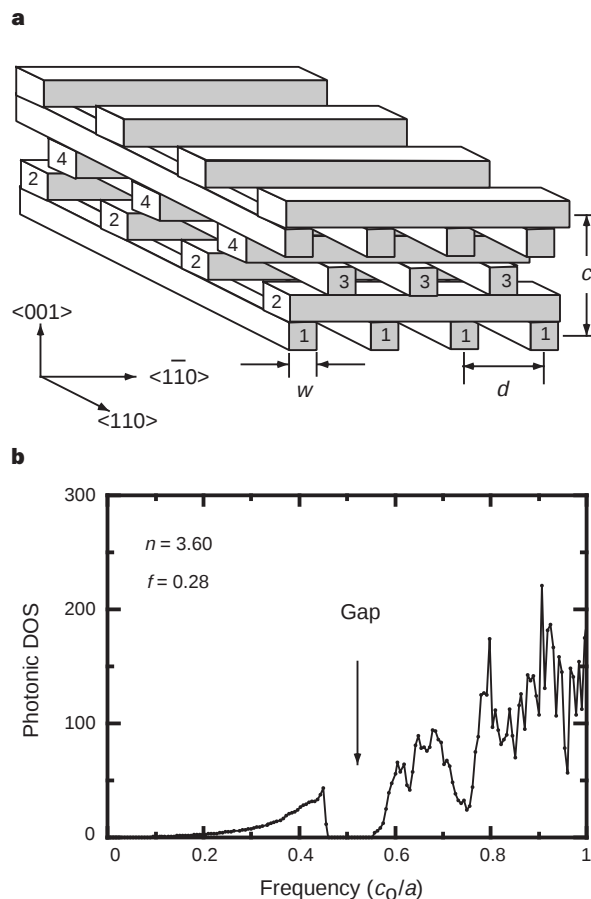


Figure 1 Structure and properties of the photonic crystal. **a**, Diagram of the layer-by-layer 3D photonic crystal. It consists of layers of one-dimensional rods with a stacking sequence that repeats itself every four layers, a unit cell, with a repeat distance of c . **b**, Computed photonic density of states (DOS) plot, assuming that the refractive index of the rods is $n = 3.60$, the filling fraction of the 3D structure $f = w/d = 0.28$, and $c/d = 1.414$. A complete photonic bandgap exists from $0.46c_0/a$ to $0.56c_0/a$, where c_0 is the speed of light *in vacuo*.

final SiO₂ removal. Figure 2a shows a scanning electron micrograph (SEM) top view of a completed four-layer structure. It has an impressive periodicity over an area of (1 cm × 1 cm). The underlying layer structure is also evident. In Fig. 2b, an SEM cross-sectional view of the same 3D photonic crystal is shown. The pitch between adjacent rods is $d = 4.2 \mu\text{m}$, the rod width is $w = 1.2 \mu\text{m}$ and the layer thickness is $1.6 \mu\text{m}$. The smoothness of the planarized surface is controlled to within 1% of the layer thickness across the entire 6-inch wafer. As the preparation of this 3D lattice involves the use of well developed and supported Si integrated-circuit fabrication process, it is suitable for large-scale integration and production.

To test the transmission property of the devices, we used a standard room-temperature Fourier-transform infrared measurement system with a broad spectral response from 2.5 to 25 μm . Before measurement, the back of the Si substrate was polished to a smoothness of better than 0.3 μm to avoid significant light scattering. The sample was cleaved to a size of 1 cm × 1 cm, which corresponds to 2,400 rods per layer, thus providing an excellent approximation of a crystal of 'infinite size' in the polycrystalline-silicon layer plane. The sample was characterized using a beam that is collimated to within a 10° divergent angle and has a spot size of ~3 mm. The infrared light is incident along the stacking direction of the 3D photonic crystal and is unpolarized. To find the absolute transmittance, a reference spectrum from a bare Si wafer was first obtained along with the signal spectrum taken from a wafer with a 3D photonic crystal built on it. By ratioing the signal to reference

spectrum, the system's detector response was normalized, thereby eliminating the small unwanted absorption and other losses in silicon substrate.

The absolute transmission spectrum of light propagating along the (001) direction of the 3D photonic crystal, that is, normal to the substrate, is shown in Fig. 3. The transmittance is plotted on a logarithmic scale as a function of wavelength from 6 to 21 μm . At 10–14 μm , a strong transmittance dip is seen, signifying the existence of a photonic bandgap in the 3D structure. The gap extends over a spectral range $\Delta\lambda = 4.5 \mu\text{m}$; the gap to mid-gap ratio $\Delta\lambda/\lambda = 40\%$ is large. The development of the bandgap is also clearly evident as the number of overlayers is increased one at a time from 2 to 5 (Fig. 3 inset). Such a large 3D gap would allow strong photon localization with the gap^{12,13}, as well as a detailed manipulation of photonic defect states^{14,15}. The predicted photonic band edge positions along the (001) orientation are shown as arrows in Fig. 3. The agreement between theory and experiment is excellent. At longer wavelengths, 16–21 μm , the transmittance stays close to the expected value of 100%. We attribute the cleanliness of the data to the device uniformity across the wafer, and to a precise control of the device critical dimensions.

From the evolution of the transmittance dip, the attenuation constant of light propagating inside the photonic crystal is derived. The inset of Fig. 3 shows the transmittance versus number of layers at three different wavelengths, $\lambda = 11 \mu\text{m}$, and corresponds to an attenuation of 12 dB per unit cell, that is, four layers. With as little as two unit cells, which is only about one wavelength thick, 99% of the incident light is reflected.

In an attempt to probe the angular dependent of the photonic bandgap, which has its minimum at a crystal orientation halfway between (001) and (110) axes, a tilt-angle transmission experiment was conducted. In the inset of Fig. 4, a diagram of the experimental configuration is shown. The light is incident along the plane spanned by the (001) and (110) vectors. The incident angle is varied from θ values of 0° to 60° in free space, which would correspond to smaller angles θ as light propagates into and through the higher-index 3D photonic crystal. The transmittance data (Fig. 4, main figure) shows a systematic narrowing of the photonic bandgap from $\Delta\lambda = 4.5 \mu\text{m}$ to $\Delta\lambda = 3 \mu\text{m}$ as θ is increased to 60°. The observed photonic bandgap is close to, but slightly larger than,

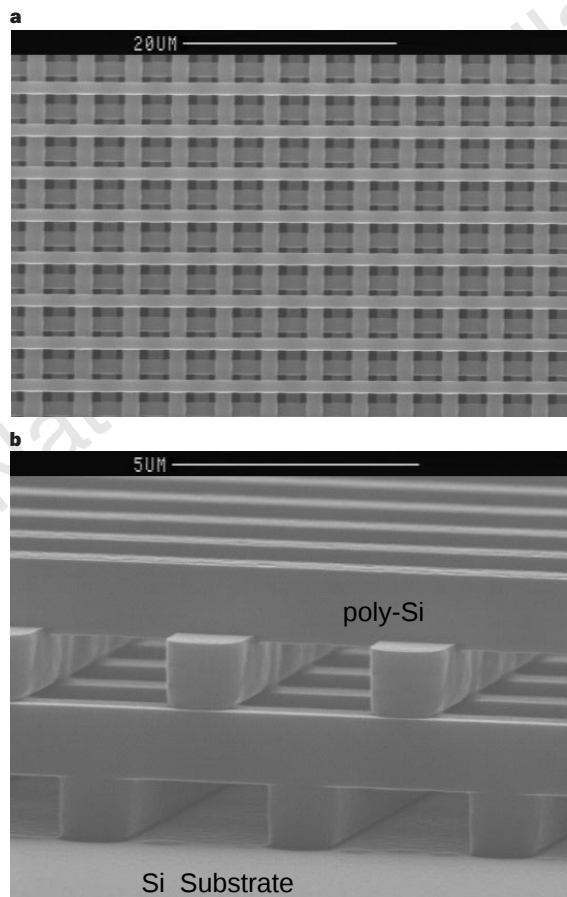


Figure 2 Micrographs of the photonic crystal. **a**, SEM top view of a completed four-layer structure. It shows good periodicity. The underlying layer structures are also evident. Scale bar, 20 μm . **b**, SEM cross-sectional view of the same 3D photonic crystal. The rods are made of polycrystalline silicon. The spacing between adjacent rods is $d = 4.2 \mu\text{m}$, the rod width is $w = 1.2 \mu\text{m}$, and the layer thickness is $1.6 \mu\text{m}$. Scale bar, 5 μm .

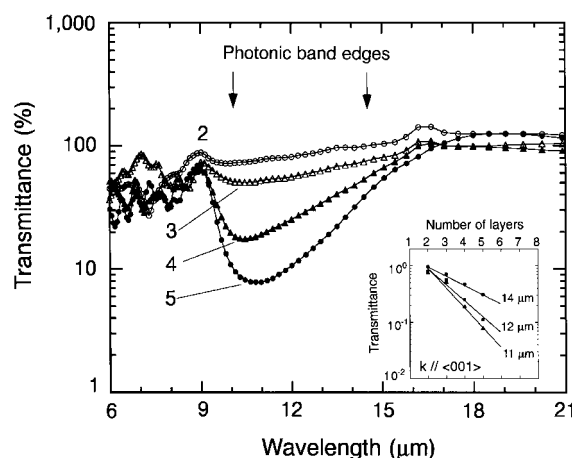


Figure 3 Transmission spectrum for the 3D photonic crystal shown in Fig. 2. The light propagates along the stacking direction, (001) and its E-field is unpolarized. At 10–14 μm , a strong dip is seen, signifying the existence of a photonic bandgap in our 3D structure. The development of the bandgap is also evident as the number of overlayers is increased from 2 to 5. Inset, semi-log plot of the transmittance versus number of layers at three different wavelengths: λ is 11, 12 or 14 μm . The strongest attenuation occurs at $\lambda = 11 \mu\text{m}$, with an attenuation strength of 12 dB per unit cell.

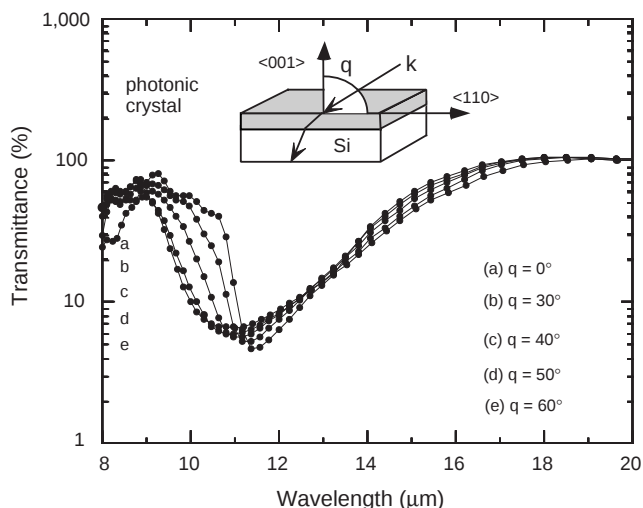


Figure 4 Tilt-angle transmission spectra taken at incident angles θ of 0°, 30°, 40°, 50° or 60° for the five-layer photonic crystal. A diagram of the experimental configuration is shown in the inset.

the expected minimum gap of $\Delta\lambda = 2.3 \mu\text{m}$, for the f.c.c. lattice. The calculated band edges from Fig. 1b are $\lambda = 10.6 \mu\text{m}$ and $12.9 \mu\text{m}$. Although the 3D crystal is not designed for a complete mapping of the photonic bandgap dispersion, the tilt-angle results are consistent with theoretical prediction that our 3D structure shows a complete photonic bandgap.

For completeness, we studied the minimum refractive-index contrast necessary for generating a photonic bandgap in a layer-by-layer 3D structure. We examined 3D photonic crystals consisting of Si/SiO₂ layers, which has a refractive index contrast of 2.4. A bandgap at $\lambda = 12.5 \mu\text{m}$ was found that extends over a $\Delta\lambda = 1 \mu\text{m}$ spectral range. By optimizing design parameters such as the dielectric filling fraction, our calculation suggests that it will be possible to accomplish a photonic bandgap at a refractive-index contrast as low as 2.0. In this case, building 3D photonic crystals from materials such as Si₃N₄ may become feasible. As this type of structure is more robust, it is potentially useful for applications that require a large area of photonic crystals.

Our experimental realization of a Si-based 3D photonic crystal in the infrared opens a door for Si photonic-crystal devices that are suitable for large-scale integration. Owing to its large photonic bandgap, a 3D photonic crystal can be used as a bandpass filter that may be integrated with, for example, a Si waveguide or a photo-detector. By creating a single-mode 3D defect cavity, a narrowband bandpass filter with an adjustable bandwidth is readily available¹⁶. Furthermore, with an attenuation constant of 12 dB per unit cell, the 3D structure is capable of producing a high-quality resonant cavity with a quality factor (Q) exceeding 10,000 and confining light to a volume that is a fraction of $(\lambda/n)^3$ (ref. 17). We note that a 3D photonic crystal could be used to modify and/or suppress the intrinsic thermal emission of a hot object: this would find important applications in the field of infrared emissivity engineering and thermal-signature recognition¹⁸. □

Received 17 February; accepted 8 April 1998.

1. Yablonovitch, E., Photonic band-gap structures. *J. Opt. Soc. Am. B* **10**, 283–295 (1993).
2. Joannopoulos, J., Meade, R. & Winn, J. *Photonic Crystals* (Princeton Univ. Press, New York, 1995).
3. Foresi, J. et al. Photonic band gap microcavities in optical waveguides. *Nature* **390**, 143–145 (1997).
4. McIntosh, K. et al. Three-dimensional metalodielectric photonic crystals exhibiting resonant infrared stop bands. *Appl. Phys. Lett.* **70**, 2937–2939 (1997).
5. Cheng, C. et al. Nanofabricated three dimensional photonic crystals operating at optical wavelengths. *Phys. Scripta T* **68**, 17–20 (1996).
6. Ho, K. et al. Photonic band gaps in three dimensions: new layer-by-layer periodic structures. *Solid State Commun.* **89**, 413–416 (1994).
7. Sozuer, H. & Dowling, J. Photonic band calculations for woodpile structures. *J. Mod. Opt.* **41**, 231–239 (1994).

8. Fan, S. et al. Design of three-dimensional crystals at submicron lengthscales. *Appl. Phys. Lett.* **65**, 1466–1468 (1994).
9. Ozbay, E. et al. Measurement of three-dimensional photonic band gap in a crystal structure made of dielectric rods. *Phys. Rev. B* **50**, 1945–1948 (1994).
10. Wanke, M. et al. Laser rapid prototyping of photonic band-gap microstructures. *Science* **275**, 1284–1286 (1997).
11. Patrick, W. et al. Applications of chemical mechanical polishing to the fabrication of VLSI circuit interconnections. *J. Electrochem. Soc.* **138**, 1778–1784 (1991).
12. Garcia, N. & Genack, A. Anomalous photon diffusion at the threshold of the Anderson localization transition. *Phys. Rev. Lett.* **66**, 1850–1853 (1991).
13. John, S. Localization of light. *Phys. Today* **May**, 32–40 (1991).
14. Lin, S. & Arjavalingam, G. Photonic bound states in a two-dimensional photonic crystals probed by coherent-microwave transient spectroscopy. *J. Opt. Soc. Am. B* **11**, 2124–2127 (1994).
15. Krauss, T. et al. Two-dimensional photonic bandgap structures operating at near-infrared wavelengths. *Nature* **383**, 699–702 (1996).
16. Lin, S. et al. Photonic band gap quantum well and quantum box structures: a high-Q resonant cavity. *Appl. Phys. Lett.* **68**, 3233–3235 (1996).
17. Ozbay, E. et al. Defect structures in a layer-by-layer photonic band-gap crystal. *Phys. Rev. B* **51**, 13961–13965 (1995).
18. Hopkins, F. Military laser applications. *Opt. Photonic News* **9**, 33–38 (1998).

Acknowledgements. We thank J. R. Wendt and G. A. Vawter for discussions, and S. Sucher, P. Shea and the rest of the Material Development Laboratory processing team for support. The work at Sandia National Laboratories is supported through the US Department of Energy. Sandia is a multiprogramme laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the US Department of Energy. Ames Laboratory is operated for the US Department of Energy by Iowa State University.

Correspondence and requests for materials should be addressed to S.Y.L. (e-mail: SLIN@sandia.gov).

An electrophoretic ink for all-printed reflective electronic displays

Barrett Comiskey, J. D. Albert, Hidekazu Yoshizawa & Joseph Jacobson

Massachusetts Institute of Technology, The Media Laboratory, 20 Ames Street, Cambridge, Massachusetts 02139-4307, USA

It has for many years been an ambition of researchers in display media to create a flexible low-cost system that is the electronic analogue of paper. In this context, microparticle-based displays^{1–5} have long intrigued researchers. Switchable contrast in such displays is achieved by the electromigration of highly scattering or absorbing microparticles (in the size range 0.1–5 μm), quite distinct from the molecular-scale properties that govern the behaviour of the more familiar liquid-crystal displays⁶. Microparticle-based displays possess intrinsic bistability, exhibit extremely low power d.c. field addressing and have demonstrated high contrast and reflectivity. These features, combined with a near-lambertian viewing characteristic, result in an ‘ink on paper’ look⁷. But such displays have to date suffered from short lifetimes and difficulty in manufacture. Here we report the synthesis of an electrophoretic ink based on the microencapsulation of an electrophoretic dispersion⁸. The use of a microencapsulated electrophoretic medium solves the lifetime issues and permits the fabrication of a bistable electronic display solely by means of printing. This system may satisfy the practical requirements of electronic paper.

Previous approaches to fabricating particle-based displays have been based on rotating bichromal spheres in glass cavities¹ or elastomeric slabs^{2,3}, and electrophoresis in glass cavities^{4,5}. The advantageous optical and electronic characteristics of microparticle displays, as compared to liquid-crystal displays, result from the ability to electrostatically migrate highly scattering pigments such as titanium dioxide ($n = 2.7$), yielding a difference in optical index of refraction with the surrounding dielectric liquid of $\Delta n = 1.3$ and black pigments with very high absorption and hiding power. This results in a very short scattering length ($\sim 1 \mu\text{m}$) and a correspondingly high reflectivity and contrast, similar to that of ink on paper. For comparison, typical index differences in a scattering-mode liquid crystal are less than $\Delta n = 0.25$ (ref. 9).

Despite these favourable attributes, microparticle displays at present suffer from a number of shortcomings. These include, in the case of the bichromal sphere system, difficulty in obtaining complete rotation (and thus complete contrast) due to a fall-off in the dipole force close to normal angles and difficulty of manufacture. In the case of electrophoretic systems, the shortcomings include reduced lifetime due to colloidal instability¹⁰. Finally, bistable display media of all types, to date, are not capable of being manufactured with simple processes such as printing. (Although recent results have been obtained in ink-jetting of electroluminescent doped polymer films for organic light-emitting structures¹¹, such emissive non-bistable displays do not meet a key criterion of 'electronic paper', namely the persistent display of information with zero power consumption).

To realize a printable bistable display system, we have synthesized an electrophoretic ink by microencapsulating droplets of an electrophoretic dispersion in individual microcapsules with diameters in the range of 30–300 μm . Figure 1a indicates schematically the operation of a system of microencapsulated differently coloured and charged microparticles, in which one or the other species of particles may be migrated towards the viewer by means of an externally applied electric field. Figure 1b shows cross-sectional photomicrographs of a single microcapsule addressed with positive and negative fields.

The microencapsulated electrophoretic system was synthesized using the following process. The internal phase of the microcapsules was composed of a mixed dispersion of black and white microparticles in a dielectric fluid. To obtain white microparticles, a suspension of rutile titanium dioxide (specific gravity = 4.2) in molten low-molecular-weight polyethylene was atomized. A similar process was used to prepare black microparticles with an inorganic black pigment. The resulting particles were sieved to obtain a dry powder with an average diameter of 5 μm . Alternatively, smaller monodisperse particles, with a diameter less than 1 μm , were prepared chemically. The polyethylene serves to reduce the specific gravity of the particles (typically to ~ 1.5) and present a modified surface chemistry for charging purposes. The particles in suspension acquire a surface charge due to the electrical double layer¹². The black and white particles have different zeta potential (and hence different mobility) due to the electroconductivity of the black pigment¹³. These microparticles are then dispersed in a mixture of

tetrachloroethylene (specific gravity, 1.6) and an aliphatic hydrocarbon (specific gravity, 0.8) which is specific-gravity-matched to the manufactured particles. In the case where the particles are designed with charges of opposite sign, they are prevented from coagulation by providing a physical polymeric adsorbed layer on the surface of each particle which provides strong inter-particle repulsive forces¹⁰. Alternatively, a single particle system (white microparticles) dispersed in a dyed (Oil Blue N) dielectric fluid was prepared.

This suspension, which in typical electrophoretic displays would be interposed between two glass electrodes, was then emulsified into an aqueous phase and microencapsulated by means of an *in situ* polycondensation of urea and formaldehyde. This process produces discrete mechanically strong and optically clear microcapsules. The microcapsules are optionally filtered to obtain a desired size range, and are subsequently washed and dried.

Microcapsules (white particle in dye) were prepared and dispersed in a carrier (ultraviolet-curable urethane) and subsequently coated onto a transparent conductive film (indium tin oxide (ITO) on polyester). Rear electrodes printed from a silver-doped polymeric ink were then applied to the display layer. Figure 2 shows a series of microphotographs at different magnifications in which the letter 'k' has been electronically addressed in the electronic ink. The sample shown was sieved to have a mean capsule size of $40 \pm 10 \mu\text{m}$ yielding a capsule resolution of ~ 600 dots per inch. By going to a system using structured top electrodes (or address lines as opposed to a continuous electrode), we were able to fractionally address single microcapsules, yielding an addressable resolution of $\sim 1,200$ dots per inch.

In addition to making a flexible and printable system, microencapsulating the electrophoretic dispersion has solved a long-standing problem with electrophoretic displays, namely limited lifetime due to particle clustering, agglomeration and lateral migration¹⁰. This is because the dispersion is physically contained in discrete compartments, and cannot move or agglomerate on a scale larger than the capsule size. In ink samples that we have prepared to date, $>10^7$ switching cycles have been observed with no degradation in performance. We have measured contrast ratios of 7:1 with 35% reflectance and a near-lambertian viewing characteristic. Using the same measurement system, we measured newspaper at 5:1 contrast and 55% reflectance. We have observed image storage

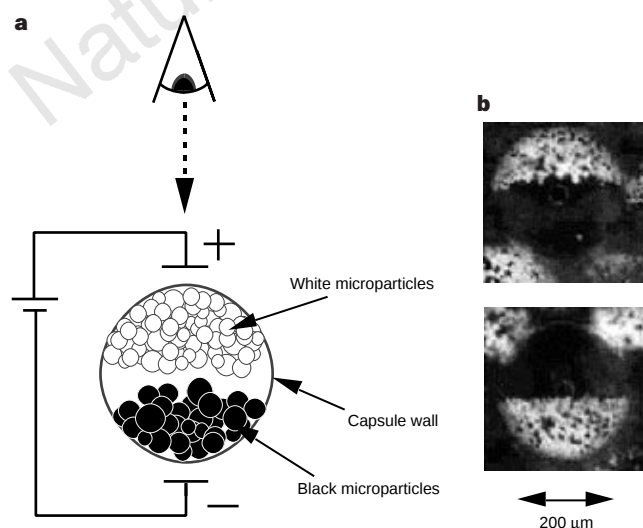


Figure 1 Electrophoretic microcapsule. **a**, Schematic illustration of microencapsulated electrophoretic image display (white and black microparticles system). The top transparent electrode becomes positively charged, resulting in negatively-charged white microparticles migrating towards it. Oppositely charged black microparticles move towards the bottom electrode. **b**, Photomicrograph of an individual microcapsule addressed with a positive and negative field.

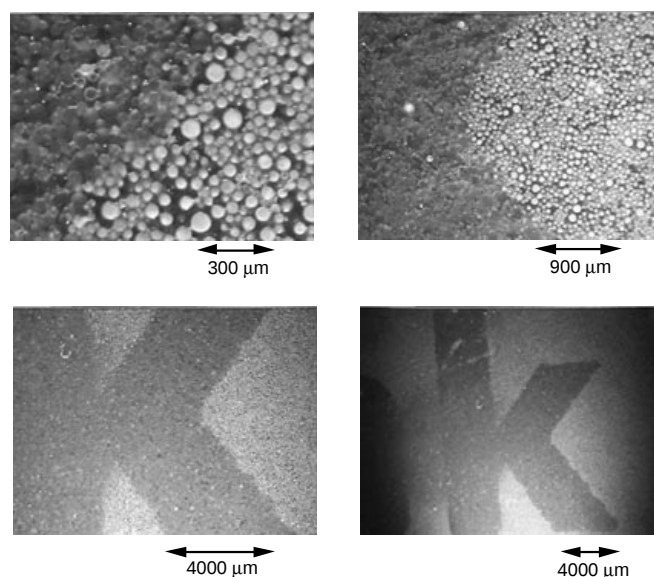


Figure 2 Electrophoretic ink. Photomicrographs of 200- μm -thick film of electronic ink ('white particles in dye' type) with a capsule diameter of $40 \pm 10 \mu\text{m}$ (top view). The electronically addressed letter 'k' is white, other areas are blue.

times of several months, after which the material may be addressed readily.

In Fig. 3a we show maximum minus minimum reflected optical power (non-normalized contrast) versus drive frequency for several different applied fields; open symbols indicate data taken from low to high frequency, filled symbols show data from high to low frequency. Figure 3b shows plots of contrast versus applied field for different frequencies. The data in Fig. 3 were obtained for a 200- μm film of electronic ink (of the type 'white particles in dye'), with capsules of diameter $40 \pm 10 \mu\text{m}$ on ITO polyester with a silver-ink rear electrode.

We may calculate the particle mobility, zeta potential and charge per particle as follows. The low-frequency asymptote in Fig. 3a indicates that at sufficiently slow driving frequency there is time for the internal particles to make a full traverse of the capsule, after which there is no further contribution to the contrast. Thus the cusp of the sigmoid in Fig. 3a may be taken to indicate the frequency (5 Hz) at which the particle just traverses a round trip ($2 \times 40 \mu\text{m} = 80 \mu\text{m}$) in the microcapsule yielding a velocity of $v = 400 \mu\text{m s}^{-1}$. The Reynolds number is given by $\text{Re} = \rho v r \eta^{-1}$ where ρ is the internal fluid density, r is the particle radius, η is the internal fluid viscosity and v is the particle velocity. For the present system we have $\rho = 1.6 \times 10^{-15} \text{ kg } \mu\text{m}^{-3}$, $v = 400 \mu\text{m s}^{-1}$, $r = 0.5 \mu\text{m}$, and $\eta = 0.8 \times 10^{-9} \text{ kg } \mu\text{m}^{-1} \text{ s}^{-1}$; and $\text{Re} = 0.0004 \ll 1$ and the flow is laminar¹⁴. The transient time to establish the laminar flow may be estimated from the Navier-Stokes equation to be $\tau_{ss} = (1/9)r^2\rho\eta^{-1} = 55 \text{ ns}$. Thus for time-

scales of interest, the particle mobility is given as $\mu = v/E = \epsilon\zeta/6\pi\eta = q/12\pi r\eta$ where ϵ is the dielectric constant of the internal fluid, ζ is the zeta potential of the particles and q is the charge per particle. Thus for our system we have $\mu = 169 \mu\text{m}^2 \text{ V}^{-1} \text{ s}^{-1}$, $\epsilon = 2.5 \times 8.85 \times 10^{-6} \text{ kg } \mu\text{m s}^{-2} \text{ V}^{-2}$, $\zeta = 120 \text{ mV}$, and $q = 2.6 \times 10^{-18} \text{ C} = 16e^-$. The very small charge per particle indicates the mechanism for 'field-off' bistability. Particle/capsule-wall and particle/particle binding dominate the particle/particle repulsion coming from the very small charge per particle. Other mechanisms for 'field-off' bistability include charge redistribution and minimization.

Figure 3a indicates the compromise between switching speed and contrast; Fig. 3b shows the compromise between applied field and contrast. In both sets of curves, the high-contrast asymptote indicates the regime in which the particles have made full traverse of the capsule, and further time or applied field has no further consequence. Two-particle (black/white particles) systems, as opposed to 'single particle in dye' systems, relax these trade-offs due to the absence of exponential light absorption in the dye case due to dye interstitially present between white particles.

In order to efficiently address a large number of pixels, matrix addressing is required. Matrix addressing in turn requires that either the contrast material (passive matrix) or an underlying electronic layer (active matrix) possess a threshold in order to prevent crosstalk between address lines. Typically, the display material may support passive matrix addressing for small numbers of pixels¹⁵. Larger numbers of pixels require the nonlinearity of an electronic element as implemented in active matrix addressing¹⁶. Active matrix structures, typically formed from polysilicon on glass, are expensive and would obviate many of the advantages of a flexible low-cost paper-like display. Our group has recently demonstrated an all-printed metal-insulator-metal (MIM) diode structure¹⁷ capable of forming the active matrix for electrophoretic ink, which should enable us to drive high-pixel-count sheets of electrophoretic ink while maintaining the ease of fabrication and low-cost structure provided by the ink material itself. Work by other groups may also be of use in this endeavour¹⁸. We believe that such systems will open up a new field of research in printable microscopic electro-mechanical systems. Coupled with recent advances in printable logic, such technology offers the prospect of fundamentally changing the nature of printing, from printing form to printing function. $\square$

Received 8 December 1997; accepted 18 May 1998.

1. Pankove, J. I. *Color Reflection Type Display Panel* (Tech. Note No. 535, RCA Lab., Princeton, NJ, 1962).
2. Sheridan, N. K. & Berkovitz, M. A. The gyron—a twisting ball display. *Proc. Soc. Information Display* **18**(3,4), 289–293 (1977).
3. Sheridan, N. K. et al. in *Proc. Int. Display Research Conf.* (ed. Jay Morreale) L82–L85 (Toronto, 1997).
4. Ota, I., Honishi, J. & Yoshiyama, M. Electrophoretic image display panel. *Proc. IEEE* **61**, 832–836 (1973).
5. Dalisa, A. L. Electrophoretic display technology. *IEEE Trans. Electron Devices* **ED-24**(7), 827–834 (1977).
6. Castellano, J. A. & Harrison, K. J. in *The Physics and Chemistry of Liquid Crystal Devices* (ed. Sprokel, G. J.) 263–288 (Plenum, New York, 1980).
7. Fitzhenry-Ritz, B. Optical properties of electrophoretic image displays. *IEEE Trans. Electron Devices* **ED-28**(6), 726–735 (1981).
8. Comiskey, B., Albert, J. D. & Jacobson, J. Electrophoretic ink: a printable display material. In *Digest of Tech. Papers* 75–76 (Soc. Information Display, Boston, 1997).
9. Okada, M., Hatano, T. & Hashimoto, K. Reflective multicolor display using cholesteric liquid crystals. In *Digest of Tech. Papers* 1019–1026 (Soc. Information Display, Boston, 1997).
10. Murau, P. & Singer, B. The understanding and elimination of some suspension instabilities in an electrophoretic display. *J. Appl. Phys.* **49**, 4820–4829 (1978).
11. Hebner, T. R., Wu, C. C., Marcy, D., Lu, M. H. & Sturm, J. C. Ink-jet printing of doped polymers for organic light emitting devices. *Appl. Phys. Lett.* **72**, 519–521 (1998).
12. Fowkes, F. M., Jinnai, H., Mostafa, M. A., Anderson, F. W. & Moore, R. J. in *Colloids and Surfaces in Reprographic Technology* (eds Hair, M. & Croucher, M. D.) (Am. Chem. Soc., Washington DC, 1982).
13. Claus, C. J. & Mayer, E. F. in *Xerography and Related Processes* (eds Dessauer, J. H. & Clark, H. E.) 341–373 (Focal, New York, 1965).
14. Landau, L. D. & Lifshitz, E. M. *Fluid Mechanics* (Pergamon, New York, 1959).
15. Ota, I., Sato, T., Tanaka, S., Yamagami, H. & Takeda, H. Electrophoretic display devices. In *Laser* **75** 145–148 (Optoelectronics Conf. Proc., Munich, 1975).
16. Shiffman, R. R. & Parker, R. H. An electrophoretic image display with internal NMOS address logic and display drivers. *Proc. Soc. Information Display* **25**, 105–115 (1984).
17. Park, J. & Jacobson, J. in *Proc. Materials Research Soc. B8.2* (Mater. Res. Soc., Warrendale, PA, 1998).
18. Service, R. F. Patterning electronics on the cheap. *Science* **278**, 383 (1997).

Correspondence and requests for materials should be addressed to J.J. (e-mail: jacobson@media.mit.edu).

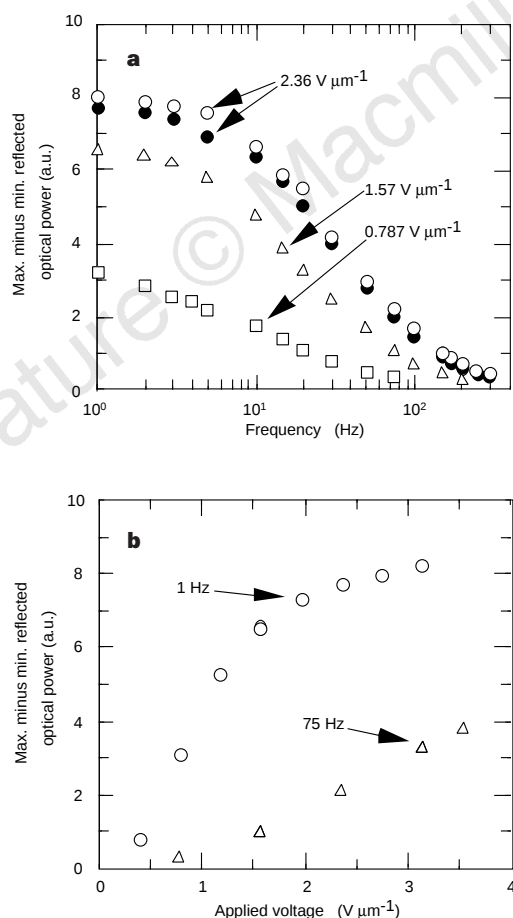


Figure 3 Properties of a 200- μm -thick film of electronic ink ('white particles in dye' type) with capsule diameter of $40 \pm 10 \mu\text{m}$. **a**, Maximum minus minimum reflected optical power (contrast) versus drive frequency for several different applied fields (open symbols, data taken from low to high frequency; filled symbols, from high to low frequency). **b**, Plot of contrast versus applied field for different frequencies.

Continuous self-assembly of organic–inorganic nanocomposite coatings that mimic nacre

Alan Sellinger*, Pilar M. Weiss*, Anh Nguyen*, Yunfeng Lu*, Roger A. Assink*, Weiliang Gong† & C. Jeffrey Brinker*

* Sandia National Laboratories and the University of New Mexico, NSF Center for Micro-Engineered Materials, The Advanced Materials Laboratory, 1001 University Blvd. SE, Albuquerque, New Mexico 87106, USA

† Department of Earth and Planetary Sciences, The University of New Mexico, Albuquerque, New Mexico 87106, USA

Nanocomposite materials are widespread in biological systems. Perhaps the most studied is the nacre of abalone shell, an orientated coating composed of alternating layers of aragonite (CaCO_3) and a biopolymer. Its laminated structure simultaneously provides strength, hardness and toughness: containing about 1 vol. % polymer, nacre is twice as hard and 1,000 times as tough as its constituent phases¹. Such remarkable properties have inspired chemists and materials scientists to develop synthetic,

'biomimetic' nanocomposite assemblies^{2–5}. Nonetheless, the efficient processing of layered organic–inorganic composites remains an elusive goal. Here we report a rapid, efficient self-assembly process for preparing nanolaminated coatings that mimic the structure of nacre. Beginning with a solution of silica, surfactant and organic monomers, we rely on evaporation during dip-coating to induce the formation of micelles and partitioning of the organic constituents into the micellar interiors⁶. Subsequent self-assembly of the silica–surfactant–monomer micellar species into lyotropic mesophases⁷ simultaneously organizes the organic and inorganic precursors into the desired nanolaminated form. Polymerization fixes this structure, completing the nanocomposite assembly process. This approach may be generalized both to other composite architectures and to other materials combinations.

Natural nanocomposites are formed by biomineralization⁵, a templated self-assembly process in which pre-organized organic surfaces regulate the nucleation, growth, morphology and orientation of inorganic crystals. Related synthetic, so-called 'biomimetic', approaches include crystallization beneath Langmuir monolayers⁸, crystallization on self-assembled monolayers^{3,9}, supramolecular self-assembly^{2,6,10} and sequential deposition¹¹. Of these, only the last two offer the ability to introduce periodic microstructural and compositional changes needed for nanocomposite assembly. With regard to nanolaminated structures, supramolecular self-assembly has resulted in the formation of lamellar (silica/surfactant) films¹² or

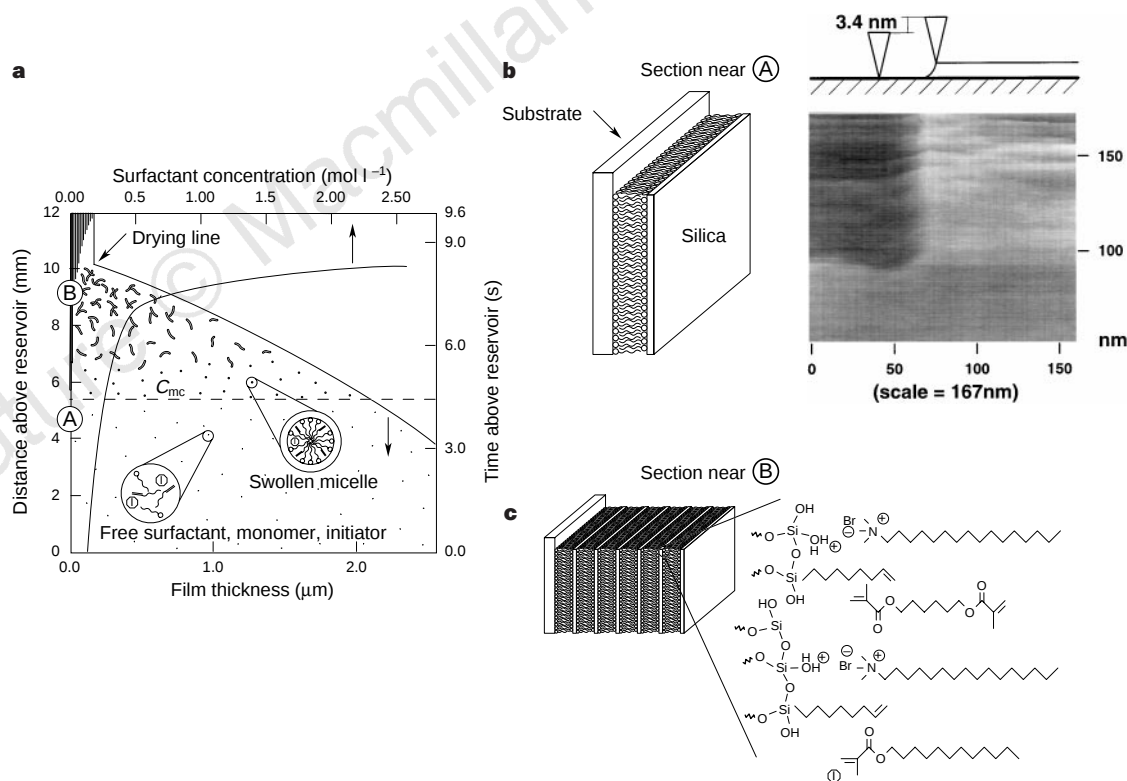


Figure 1 Diagrams showing evolution of nanolaminate structure during dip-coating (dip-coating rates ranged from 7.6 to 50.9 cm min⁻¹). **a**, Steady-state film-thinning profile established by evaporation, with vertical axes representing distance/time above sol reservoir surface, and horizontal axes showing film thickness and surfactant concentration. Preferential alcohol evaporation²⁵ progressively increases the surfactant concentration, inducing micellization and concurrent incorporation of monomer and initiator (I) into the micellar interior. **b**, Section near A. Surfactant bilayer formation occurs below C_{mc} and provides an organized surface for subsequent cooperative assembly of the lamellar mesophase. Bilayer structure was imaged using a non-contact AFM technique²⁶ following equilibration (1 h) of the silicon substrate with a silica–surfactant sol

prepared with $c_0 = 5$ wt %, but without addition of organic monomer/initiator. Step height created by scraping the bilayer with the AFM tip is consistent with a CTAB bilayer²⁷. **c**, Section near B showing the nanolaminate construction and hypothetical arrangement of surfactant, monomer, crosslinker and initiator adjacent to the oligomeric silicate surface. The chemical and structural characteristics of the coupling agent, monomers and initiators influence the local effective packing parameter of the surfactant and hence the organization of the hybrid mesophase²⁸. In the present study it is possible that the organic monomers increase the effective volume of the surfactant chains stabilizing the lamellar mesophase²⁸.

multi-bilayer composite films¹³, but such non-covalently bonded structures are mechanically unstable unless pillared¹⁴ (for example, lamellar silica-surfactant coatings collapse to amorphous silica on surfactant removal¹²). Sequential deposition has been used to prepare stable inorganic/organic nanocomposites¹¹, but this process requires many repeated deposition steps to build up a practical coating thickness.

Our efficient assembly process for a poly(dodecylmethacrylate) (PDM)/silica nanocomposite is based on a simple dip-coating procedure⁶ (see Fig. 1). The process starts with a homogeneous solution of soluble silicates, coupling agent, surfactant, organic monomers, and initiators prepared in ethanol/water solvent with

an initial surfactant concentration (c_0) below the critical micelle concentration (c_{mc}). During dip-coating, preferential evaporation of ethanol progressively enriches the concentrations of water, HCl and the non-volatile solution constituents within the depositing film. Previous optical-probe experiments performed during dip-coating of related silica/surfactant/ethanol/water sols⁶ indicated that the increasing concentrations of surfactant and water cause c to exceed c_{mc} , resulting in micelle formation and concurrent incorporation of hydrophobic optical probes into the micellar interiors. This evaporation-induced partitioning mechanism is used in the present study to incorporate alcohol soluble monomer(s) and initiators into micellar species. Continued evaporation promotes cooperative assembly of these (silica-surfactant-monomer) micellar species into interfacially organized liquid-crystalline mesophases, thereby simultaneously organizing both the inorganic and organic precursors into the desired laminated structure in a rapid (~ 10 s), continuous process. Organic polymerization (induced by light or heat), combined with continued inorganic polymerization, lock-in the nanocomposite architecture and covalently bond the organic-inorganic interface. Through variation of the surfactant type or its initial concentration, c_0 , we can exploit supramolecular self-assembly to arrive at other nanocomposite constructions.

Precursor solutions were prepared by addition of cationic surfactant, organic monomer, crosslinker, initiator and unsaturated alkyltrialkoxysilane¹⁵ (used in our study as a coupling agent) to an acidic silica sol (A2**). The acid concentration used in the A2** synthesis procedure was chosen to minimize the siloxane condensation rate¹⁶, thereby promoting facile silica-surfactant supramolecular self-assembly during film deposition⁶. In a typical preparation, TEOS ($\text{Si}(\text{OCH}_2\text{CH}_3)_4$), ethanol (EtOH), water and dilute HCl (mole ratios: 1 : 3.8 : 1 : 5×10^{-5}) were heated at 60 °C for 90 min. The sol was diluted with ethanol (1 : 2) followed by addition of water and dilute HCl. Coupling agent (7-octenyltrimethoxysilane, OTS, or methacryloxypropyltrimethoxysilane, MPS) was added followed by surfactant (cetyltrimethylammonium bromide, CTAB), monomer (dodecylmethacrylate, DM), crosslinker (hexanedioldimethacrylate, HDM), and initiator (ultraviolet,

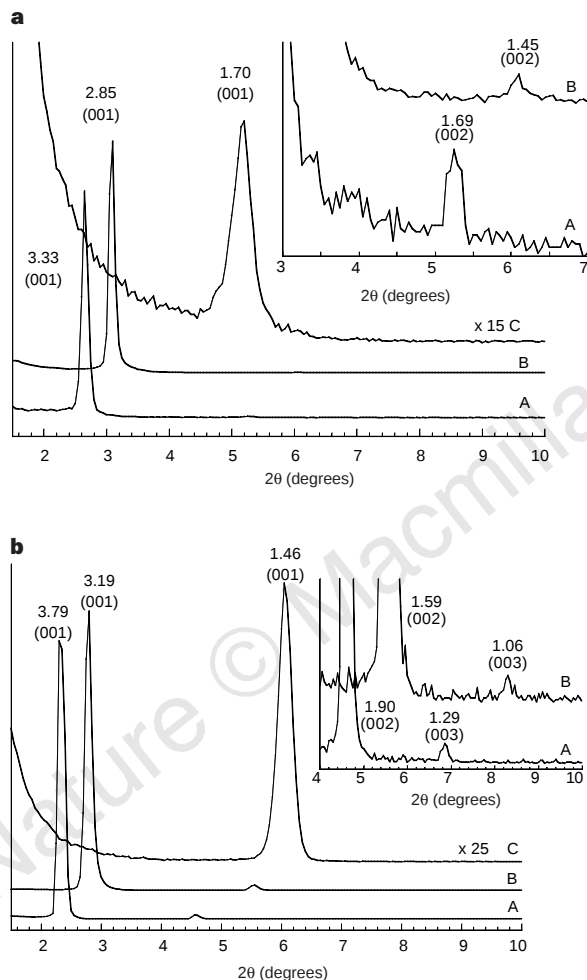


Figure 2 Thin film X-ray diffraction patterns of UV-initiated (**a**) and thermally initiated (**b**) polymerization systems. Patterns were recorded on a Siemens D500 diffractometer using Ni-filtered $\text{CuK}\alpha$ radiation with $\lambda = 1.5418 \text{ \AA}$ in θ - 2θ scan mode. Traces A, as-deposited (unpolymerized); traces B, polymerized; and traces C, polymerized and washed. Traces A and B in panels **a** and **b** contain second-order reflections indicative of a lamellar liquid crystalline mesophase. **a**, On UV irradiation of A (Hg arc lamp source with filter providing wavelengths 260–320 nm and 20 mW cm^{-2} power density), a reduction in basal cell dimension of $\sim 14\%$ is observed (trace B), consistent with shrinkage in methacrylate-based polymerization systems²⁹. An additional shrinkage of $\sim 40\%$ is observed following surfactant removal by washing in ethanol, acetone and diethylether (trace C). **b**, Thermal treatments of A (120°C for 3 h) provide a reduction in basal cell dimension of $\sim 16\%$ (B), with an additional shrinkage of $\sim 54\%$ following surfactant removal (trace C). The increased shrinkage during polymerization and washing in the thermally initiated system (**b**) versus the UV-initiated system (**a**) results from the contribution of thermally induced silica polymerization and partial volatilization of monomer during thermal treatments.

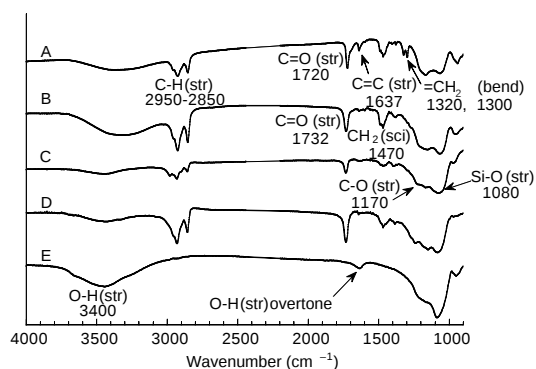


Figure 3 Fourier-transform infrared spectroscopy (FTIR) of nanocomposite coatings at various stages. Trace A, as-prepared (unpolymerized); trace B, polymerized (UV-irradiated); trace C, polymerized and washed; trace D, polymerized, washed, and treated in ammonium bifluoride to selectively etch the silica lamellae, and trace E, as-prepared and washed (without polymerization). Spectrum D contains all the bands of the bulk crosslinked PDM polymer plus siloxane bands due to the tri-functional siloxanes (coupling agents) that covalently bond the organic/inorganic interface and render the interfacial siloxane insoluble in NH_4HF_2 . The importance of polymerization to lock in the structure is demonstrated in trace E where all monomer is lost if washed before UV irradiation. (Coatings were prepared on Petri dishes. Powders scraped from the substrate surface were dispersed in KBr pellets for IR analysis. Similar results were obtained from powders scraped off silicon substrates).

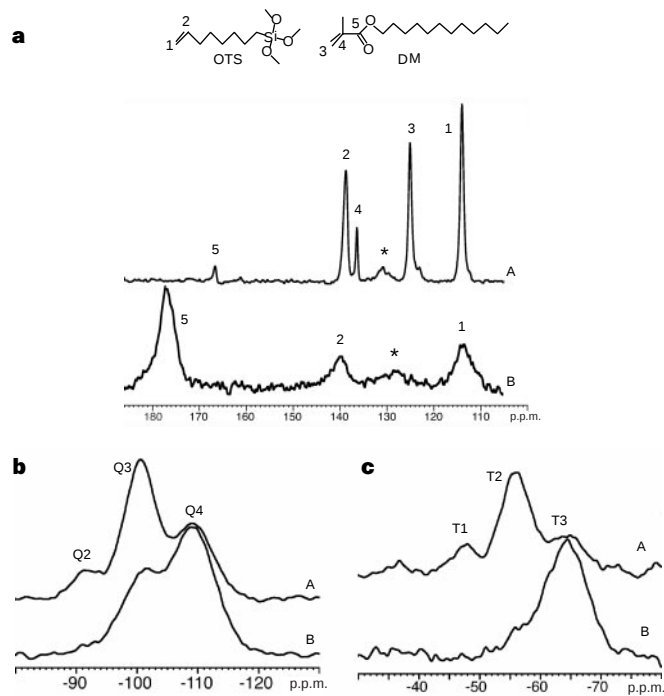


Figure 4 Magic-angle-spinning (MAS) solid-state NMR spectra of nanocomposite coatings. Trace A corresponds to as-prepared (unpolymerized), and B to polymerized (UV-irradiated), NH_3 -treated, and washed samples. **a**, ^{13}C cross-polarization (CP) MAS spectra (100.6 MHz, 2 ms CP time, 1,024 scans) of the sp^2 hybridized region showing the disappearance of vinyl and methacrylate moieties of OTS and DM following polymerization (B). This demonstrates direct covalent linkages of DM to silica via vinyl-methacrylate group co-polymerization. The presence of residual vinyl resonances in B is not surprising as their reactivity is much less than that of methacrylates in this system. Residual peaks labelled with an asterisk probably correspond to DM polymer chain end groups derived from the aromatic carbons of the photo-initiator. **b**, ^{29}Si direct polarization MAS spectra (79.5 MHz, 240 s delay time, 128 scans) with extent of reaction values for A and B of 81.5% and 90.3%, respectively, calculated by deconvolution of the Q resonances (Q2/Q3/Q4 ratios change from 0.34/1.86/1.00 to 0/0.87/1.00). **c**, ^{29}Si CP-MAS spectra (79.5 MHz, 5 ms CP time, 1,024 scans) of the T region with extent of reaction values 68.2% and 95.5%, respectively, for A and B. Qn and Tn are tetrafunctional and trifunctional silanes, respectively, where n is a numeral corresponding to the number of bridging $\text{Si}(\text{OSi})_n$ groups surrounding the Si nucleus of interest. Samples prepared as in Fig. 3. Similar ^{13}C and ^{29}Si NMR results were obtained from the thermally initiated polymerization system.

benzoin dimethylether, BME; thermal, 1,1'-azobis(1-cyclohexane-carbonitrile, ACHN)). The final reactant mole ratios were: 1 TEOS:22 EtOH:5 H_2O :0.004 HCl:0.21 surfactant:0.16 DM:0.02 HDM:0.08 OTS:0.02 initiator. All materials are based on this formulation unless noted.

Coatings were deposited on polished, (100)-silicon by dip-coating (see Fig. 1). Thicker coatings required for spectroscopic measurements were prepared by dispensing a thin layer of sol in a Petri dish followed by immediate vertical draining. After deposition, the coatings were either heated or irradiated with ultraviolet light to initiate organic polymerization. Short exposures to ammonia vapour were used to promote further condensation of the silica framework. Polymerized coatings were washed sequentially with ethanol, acetone, and diethyl ether to remove surfactant and any unpolymerized species.

Figure 2 shows a typical series of X-ray diffraction (XRD) patterns corresponding to as-deposited (unpolymerized), polymerized and washed coatings for ultraviolet (Fig. 2a) and thermally-cured (Fig. 2b) systems. The XRD pattern of the as-deposited coating in Fig. 2a

is consistent with a (001)-orientated lamellar phase with basal spacing $c = 3.33$ nm. The lamellar mesostructure is maintained during polymerization and washing, with associated basal spacing reductions of $c = 2.85$ and 1.70 nm, respectively. A similar trend is observed for the thermally initiated system. The first stage of shrinkage is due to both organic and inorganic polymerization (see following discussion) while that due to washing is a consequence of surfactant removal.

Figure 3 shows a series of Fourier-transform infrared (FTIR) spectra corresponding to the successive stages of nanocomposite formation for the ultraviolet-initiated polymerization system. Three features are observed that provide evidence of organic polymerization within the nanocomposite. (1) The $\text{C}=\text{C}$ stretch ($1,637\text{ cm}^{-1}$) of the monomer and OTS virtually disappears after ultraviolet exposure, indicating $\text{C}=\text{C} \rightarrow \text{C}-\text{C}$ conversion. (2) The $\text{C}=\text{O}$ stretching vibration in the unpolymerized film is shifted from $1,720$ to $1,732\text{ cm}^{-1}$ after ultraviolet exposure, consistent with methacrylate polymerization¹⁷ (conjugated versus unconjugated $\text{C}=\text{O}$ stretch). (3) The bandwidth of the $\text{C}=\text{O}$ -stretching peak in the polymerized film is narrower than that of low-molecular-weight polydodecylmethacrylate (PDM), suggesting that polymerization occurs within the confined geometry of the interlamellar galleries. Spectrum D in Fig. 3 is that of the polymer residue obtained after selectively etching the silica lamellae in ammonium bifluoride (NH_4HF_2). For wavenumbers greater than $\sim 1,350\text{ cm}^{-1}$, the spectrum is practically identical to that of the corresponding bulk crosslinked PDM. The presence of the $\text{Si}-\text{O}$ stretching band in spectrum D is due to tri-functional siloxanes (T) (as confirmed by ^{29}Si magic-angle spinning (MAS) NMR that covalently bond the inorganic/organic interface, rendering it insoluble in NH_4HF_2).

Further support for organic polymerization is found from ^{13}C NMR. As shown in Fig. 4a, the resonances at ~ 114 – 140 p.p.m. assigned to sp^2 hybridized carbons of $\text{C}(\text{CH}_3)=\text{CH}_2$ (DM) and $\text{CH}=\text{CH}_2$ (OTS) virtually disappear through conversion to sp^3 hybridization, and the $\text{C}=\text{O}$ resonance is shifted from 167 to 178 p.p.m., both indicative of methacrylate polymerization.

Evidence of inorganic polymerization is obtained from FTIR (see Fig. 3) and ^{29}Si MAS NMR (Fig. 4b, c). Integration of the envelope of peaks corresponding to the resonances of Q2, Q3 and Q4 silicon species shows that ammonia exposure causes the overall extent of siloxane condensation to increase from 80.2% to 88.4% and the extent of trisiloxane condensation (associated with the OTS coupling agent) to increase from 68.2% to 95.5% .

Nano-indentation measurements performed on PDM/silica, poly(4-methylstyrene)/silica and poly(4-vinylbenzylchloride)/silica nanolaminates prepared with ~ 50 wt% polymer show a three to seven times increase in indentation hardness (from 0.1 – 0.4 GPa to 0.8 – 1.0 GPa) due to combined organic/inorganic polymerization. We note that 1 GPa is the indentation hardness measured for rather dense sol-gel silica films.

The electron micrographs shown in Fig. 5 illustrate the diversity of structures attainable by this process. Figure 5a shows a transmission electron microscopy (TEM) image of a cross-section of the nanolaminated composite structure. Similar to shell, we observe a highly c -axis-orientated coating composed of successive layers of inorganic and organic polymers. Figure 5b shows a corresponding plan-view scanning electron microscopy (SEM) image, indicating that the nanocomposite coatings are featureless on micrometre length scales. Figure 5c shows a swirling pattern of organized tubules typical of hexagonal mesophases^{2,18}, and Fig. 5d shows a portion of a hierarchical composite coating composed of an isotropic worm-like micellar overlayer bonded to an orientated, nano-laminated underlayer.

Figure 5d provides support for formation mechanism depicted in Fig. 1. We propose that the lamellar structure originates from cooperative assembly of the swollen silica-surfactant micellar species with interfacially organized bilayers that form at the

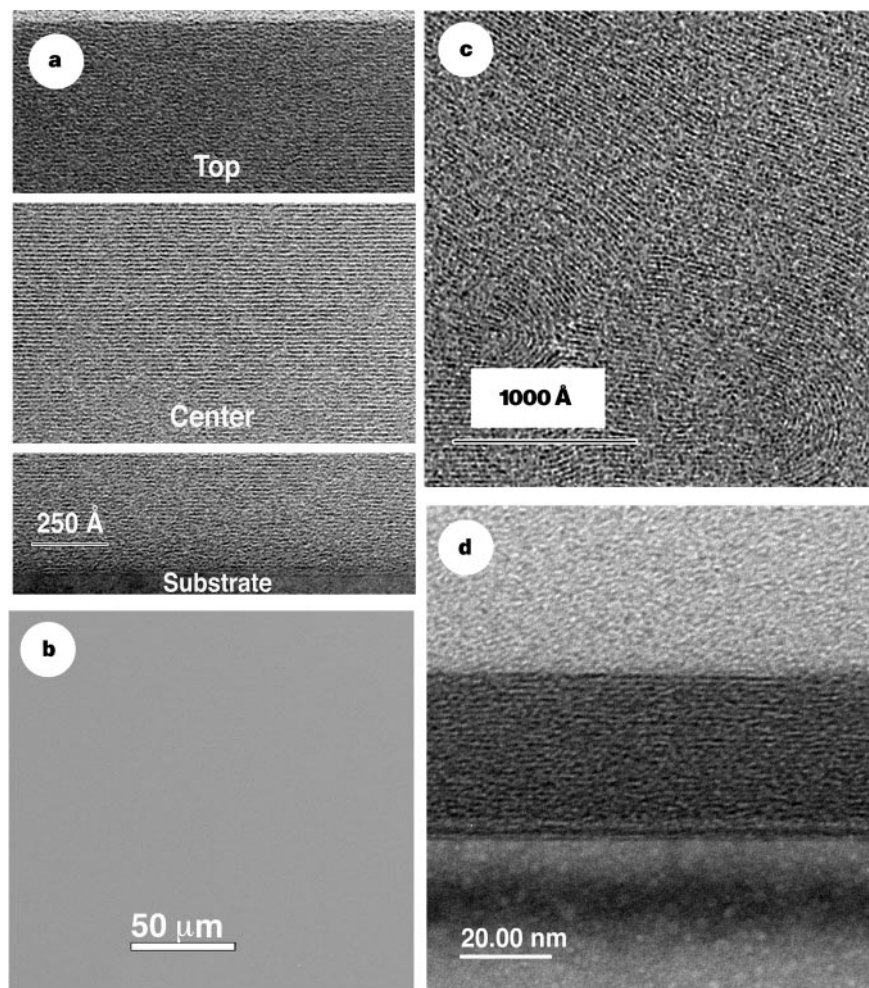


Figure 5 Electron micrographs of nanocomposite coatings prepared using the thermally initiated polymerization system ($c_0 = 5.0$ wt% CTAB) but with different coating rates or substrate surface treatments. All coatings were washed to remove surfactant and any unpolymerized species before sample preparation. **a**, TEM cross-section of ~ 500 -nm-thick film showing c -axis-orientated nanolaminate structure. Shown are representative sections of the coating in the central region and regions adjacent to the substrate and vapour surfaces. (Coating prepared on KOH-treated silicon [100] substrate at withdrawal rate of 25.4 cm min^{-1}). **b**, SEM plan-view image of nanocomposite coating prepared as in **a**. All films appear featureless on this scale. **c**, TEM image of swirling pattern of

tubules typical of hexagonal silica-surfactant mesophases formed at liquid-vapour or solid-liquid interfaces². (Coating prepared as in **a** but at withdrawal rate of 12.7 cm min^{-1}). **d**, Cross-section of coating showing interface between ordered nanolaminate structure adjacent to silicon substrate and isotropic worm-like mesostructure. During film deposition, ordered lamellar regions appear to 'grow' from the substrate by a continuous, interfacially assisted self-assembly process^{2,30}. The thickness of the ordered region depends on the process time-scale (established by the coating rate) and the chemical nature of the substrate surface. (Film prepared as in **a** but with CTAB-treated surface).

solid-liquid interface at an earlier stage of the deposition process (as observed by atomic force microscopy (AFM), Fig. 1b). An ordered lamellar mesophase 'grows' from the substrate surface by continuous accretion. The thickness of the ordered region depends on c_0 , the timescale of the dipping operation, and the chemical nature of the substrate surface.

Compared to sequential-deposition processes, the evaporation-induced partitioning and self-assembly inherent to our process allow simultaneous organization of the organic and inorganic precursors, so that thousands of layers form at once. In contrast to other nanocomposite self-assembly processes^{19,21} we covalently link the organic and inorganic interface, and the continuous nature of our dip-coating process allows rapid formation of optically transparent coatings suitable for applications such as automotive finishes, hard coats, and optical hosts. The evaporation-induced partitioning approach is generally applicable to a wide range of organic and inorganic precursors, enabling the extension of this process to other materials combinations, such as inorganic/metal²² or inorganic/conductive-polymer²³, of interest for high-capacitance devices, catalysis and quantum optics or electronics²⁴.

Received 19 February; accepted 13 May 1998.

1. Jackson, A., Vincent, J. & Turner, R. The mechanical design of nacre. *Proc. R. Soc. Lond. B* **234**, 415–425 (1988).
2. Aksay, I. *et al.* Biomimetic pathways for assembling inorganic thin films. *Science* **273**, 892–898 (1996).
3. Bunker, B. *et al.* Ceramic thin-film formation on functionalized interfaces through biomimetic processing. *Science* **264**, 48–55 (1994).
4. Heuer, A. *et al.* Innovative materials processing strategies: a biomimetic approach. *Science* **255**, 1098–1105 (1992).
5. Mann, S. Molecular tectonics in biomineralization and biomimetics materials chemistry. *Nature* **365**, 499–505 (1993).
6. Lu, Y. *et al.* Continuous formation of supported cubic and hexagonal mesoporous films by sol-gel dip-coating. *Nature* **389**, 364–368 (1997).
7. Kresge, C., Leonowicz, M., Roth, W., Vartuli, C. & Beck, J. Ordered mesoporous molecular sieves synthesized by a liquid-crystal template mechanism. *Nature* **359**, 710–712 (1992).
8. Heywood, B. & Mann, S. Template-directed nucleation and growth of inorganic materials. *Adv. Mater.* **6**, 9–19 (1994).
9. Tarasevich, B., Rieke, P. & Liu, J. Nucleation and growth of oriented ceramic films onto organic interfaces. *Chem. Mater.* **8**, 292–300 (1996).
10. Yang, H., Kuperman, A., Coombs, N., Mamiche-Afara, S., Ozin, G. Synthesis of oriented films of mesoporous silica on mica. *Nature* **379**, 703–705 (1996).
11. Keller, S., Kim, H. & Mallouk, T. Layer-by-layer assembly of intercalation compounds and heterostructures on surfaces: toward molecular "beaker" epitaxy. *J. Am. Chem. Soc.* **116**, 8817–8818 (1994).
12. Ogawa, M. Formation of novel oriented transparent films of layered silica-surfactant nanocomposites. *J. Am. Chem. Soc.* **116**, 7941–7942 (1994).
13. Okada, H., Sakata, K. & Kunitake, T. Formation of oriented iron oxide particles in cast multilayer films. *Chem. Mater.* **2**, 91–93 (1990).

14. Tanev, P. & Pinnavaia, T. Biomimetic templating of porous lamellar silicas by vesicular surfactant assemblies. *Science* **271** 1267–1269 (1996).
15. Burkett, S., Sims, S. & Mann, S. Synthesis of hybrid inorganic-organic mesoporous silica by co-condensation of siloxane and organosiloxane precursors. *Chem. Commun.* **11**, 1367–1368 (1996).
16. Brinker, C., Sehgal, R., Raman, N., Schunk, P. & Headley, T. Polymer approach to supported silica membranes. *J. Sol-Gel Sci. Technol.* **2**, 469–476 (1994).
17. Pretsch, E., Siebl, J., Simon, W. & Clerc, T. in *Tables of Spectral Data for Structure Determination of Organic Compounds* (eds Pretsch, E., Siebl, J., Simon, W. & Clerc, T.) 1141–1142 (Springer, Berlin, 1989).
18. Yang, H., Coombs, N., Dag, O., Sokolov, I. & Ozin, G. Free-standing mesoporous silica films; morphogenesis of channel and surface patterns. *J. Mater. Chem.* **7**, 1755–1761 (1997).
19. Giannelis, E. Polymer layered silicate nanocomposites. *Adv. Mater.* **8**, 29–35 (1996).
20. Gray, D., Hu, S., Juang, E. & Gin, D. Highly ordered polymer-inorganic nanocomposites via monomer self-assembly: in situ condensation approach. *Adv. Mater.* **9**, 731–736 (1997).
21. Templin, M. et al. Organically-modified aluminosilicate mesostructures from block-copolymer phases. *Science* **278**, 1795–1798 (1997).
22. Huber, C., Moller, K. & Bein, T. Reactivity of a trimethylstannyl molybdenum complex in mesoporous MCM-41. *J. Chem. Soc.* 2619–2620 (1994).
23. Wu, C. & Bein, T. Conducting polyaniline filaments in a mesoporous channel host. *Science* **264**, 1757–1759 (1994).
24. Ozin, G. Nanochemistry: synthesis in diminishing dimensions. *Adv. Mater.* **4**, 612–649 (1992).
25. Nishida, E. et al. In situ fluorescence probing of the chemical changes during sol-gel thin film formation. *J. Am. Ceram. Soc.* **78**, 1640–1648 (1995).
26. Manne, S., Cleveland, J., Gaub, H., Stucky, G. & Hansma, P. Direct visualization of the surfactant hemimicelles by force microscopy of the electrical double layer. *Langmuir* **10**, 4409–4413 (1994).
27. Bull, L. et al. in *Zeolites and Related Microporous Materials: State of the Art 1994* (eds Weitkamp, J., Karge, H., Pfeifer, H. & Holderich, W.) 429–434 (Elsevier, Amsterdam, 1994).
28. Israelachvili, J. *Intermolecular and Surface Forces* (Academic, San Diego, 1992).
29. Odian, G. *Principles of Polymerization* (Wiley, New York, 1981).
30. Yang, H., Coombs, N., Sokolov, I. & Ozin, G. Free-standing and oriented mesoporous silica films grown at the air-water interface. *Nature* **381**, 589–592 (1996).

Acknowledgements. We thank S. Singh, R. Tissot, S. Rao, Y. Guo and D. Schmale for assistance with experimental portions of this work, and M. Aragon and K. Burns for technical illustrations. This work was supported in part by the UNM/NSF Center for Micro-Engineered Materials, DOE Basic Energy Sciences, and SNL's Laboratory-Directed Research and Development Program (LDRD); and by the US Department of Energy. Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the US Department of Energy.

Correspondence and requests for materials should be addressed to C.J.B. (e-mail: cjbrink@sandia.gov).

Rapid eruption of Skye lavas inferred from precise U–Pb and Ar–Ar dating of the Rum and Cuillin plutonic complexes

M. A. Hamilton*, D. G. Pearson†, R. N. Thompson†, S. P. Kelley‡ & C. H. Emeleus†

* Continental Geoscience Division, Geological Survey of Canada, Ontario, Canada K1A 0E8

† Department of Geological Sciences, Durham University, South Road, Durham DH1 3LE, UK

‡ Department of Earth Sciences, The Open University, Milton Keynes, MK7 6AA, UK

The interpretation of rocks of the British Tertiary Volcanic Province has played an important role in the historical development of many concepts in igneous petrology. Exposures of lavas, sub-volcanic rocks and plutonic complexes have allowed a detailed understanding of the field relationships between such units in the context of flood-basalt magmatism^{1–3}. Nevertheless, age control has been a source of much controversy and a limiting factor in comparing these relationships to recent developments in the theoretical modelling of magmatism within continents⁴. Here we report precise ²⁰⁶Pb/²³⁸U zircon ages of 60.53 ± 0.08 Myr (2σ) for the Rum basic/ultrabasic pluton and 58.91 ± 0.07 Myr for the Cuillin gabbros, Skye, which tightly constrain eruption of the greater than 1.5-km-thick Skye lavas to a maximum duration of 1.6 ± 0.2 Myr. These dates yield magma production rates for the Skye lavas of about 2.2×10^{-3} km³ yr⁻¹ comparable with rates inferred for individual magmatic centres produced by melting related to mantle plumes below ocean basins. In addition, the approximately 30 km of lithospheric thinning suggested by magma chemistry is required to have occurred in less than 2 Myr.

The >1.5-km-thick Skye main lava succession (SMLS)^{5–7} is the

remnant of an important surface expression of the magmatism related to rifting of the northeast Atlantic continental margin due to impingement of the Iceland mantle plume 63 Myr ago^{8,9}. Geochemical studies on Skye lavas show that they are dominated by transitional to alkali-basalts of the SMLS^{2,10}. A later, tholeiitic magma type, Preshal-More basalt, was produced in large volumes but is now mostly preserved as dykes¹¹. Flows of Preshal-More basalts overlie the SMLS and have compositions more akin to mid-ocean-ridge basalt (MORB). The temporal change in major- and trace-element composition between the SMLS and Preshal-More magma types¹² has been interpreted in terms of progressive lithospheric thinning beneath Skye¹³. Determination of the timespan of lava eruptions will

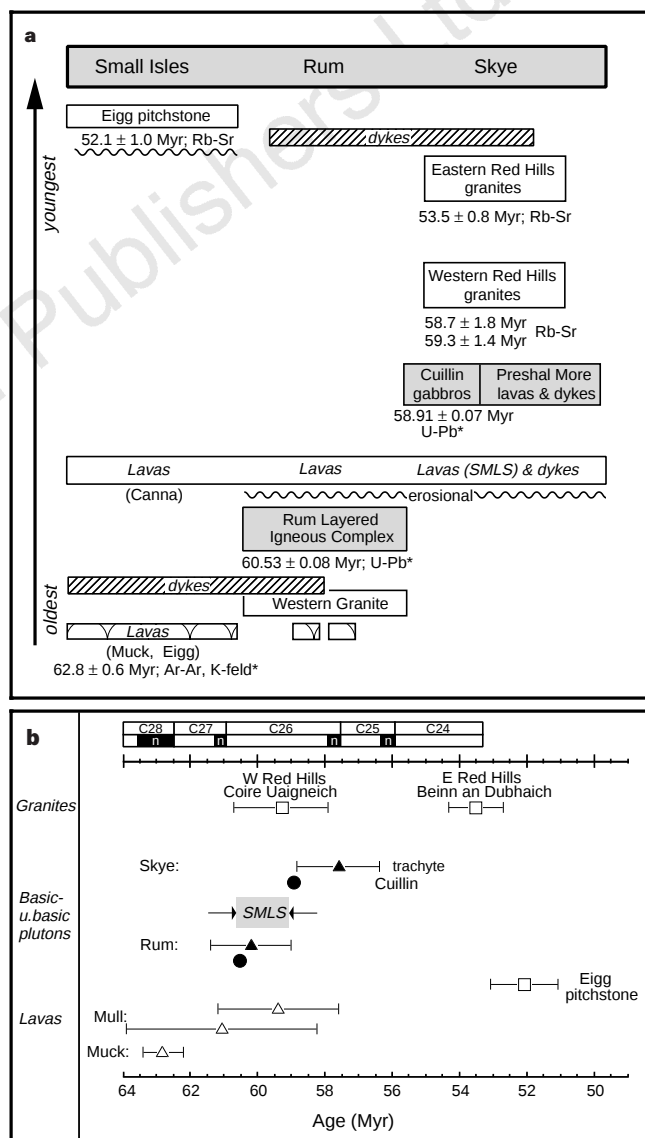


Figure 1 Field and chronostratigraphy of magmatism on Rum, Skye and the Small Isles. **a**, Diagrammatic representation of simplified igneous stratigraphical relationships established on the basis of field relations (see ref. 3 for summary). Radioisotopic ages are given with method; asterisk shows new dates; shown are data from this work and ref. 8; errors are 2σ . For a more comprehensive summary of ages and their source see ref. 9. **b**, Comparison of Rum-Skye basic/ultrabasic complex U-Pb zircon ages with other ages for Skye igneous events and Mull, Muck and Eigg lava suites. Squares indicate Rb-Sr isochrons, triangles are ⁴⁰Ar/³⁹Ar step heating ages, and circles are U-Pb zircon ages; filled symbols are data from this study. For sources of dates see Table 1 and refs 3, 8, 9, 17, 25. Arrows indicate 2σ uncertainties in all cases. Chronostratigraphic timescale is from ref. 26. Maximum likely duration of Skye lavas (SMLS) shown by shading.

thus allow estimation of the rate of lithospheric thinning beneath this part of the Hebrides, in addition to constraining the rate of magma production within the plume. This study provides high-precision absolute radioisotopic ages for the Rum and Cuillin (Skye) plutonic complexes. We then use well documented field relationships between the plutons and lavas³ to constrain the time-scale of lava eruption.

The stratigraphy³ of Tertiary magmatic activity on the Small Isles and Skye (Fig. 1) show that the relative timing of eruption of the SMLS and lavas on northwest Rum and Canna can be well constrained by the presence of eroded clasts of all members of the Rum central complex within intra-lava sediments of these successions^{3,14}. This indicates that the Rum central complex had been intruded, unroofed and eroded before the eruption of the SMLS. Dating the Rum central complex therefore provides a lower bound on the age of the Skye lava field. The earliest Cuillin magmas have SMLS compositions, but studies of isotopic and elemental compositions clearly show that the bulk of the complex is Preshal-More tholeiite¹⁵. Dykes of the main Skye swarm cut the SMLS and are predominantly Preshal-More type¹¹, injected laterally from the Cuillin centre¹⁶. This places the Cuillin centre as the likely source of both Preshal-More lavas and dykes. Consequently, final quenching of the Cuillin plutons marked the end of the main basaltic magmatic episode. Thus, dating the Cuillin complex imposes a minimum age on the main Skye basaltic magmatism.

Extreme fractional crystallization products within the Rum and Skye basic/ultrabasic plutons contain accessory phases, such as zircon and phlogopite, that facilitate precise dating. Zircons were recovered from a 15-cm-thick alkaline segregation within the margin of the Rum layered complex, and dated by isotope dilution U–Pb analysis (Table 1). Mica was also extracted from a pegmatite

at Harris Bay, Rum, and analysed using the laser $^{40}\text{Ar}/^{39}\text{Ar}$ technique. On Skye, gabbros of the Outer Layered Series of the Cuillin intrusion on Skye locally develop quartzofeldspathic pegmatite veins (1 to >20 cm thick), some containing abundant zircons. Biotites from a trachyte dyke cutting the upper portions of the main Skye lava pile were also dated by the $^{40}\text{Ar}/^{39}\text{Ar}$ method.

The age of the Rum alkaline segregation is tightly constrained by three concordant zircon fractions, which define a weighted mean $^{206}\text{Pb}/^{238}\text{U}$ age of 60.53 ± 0.08 Myr (2σ ; mean squared weighted deviates, MSWD = 0.46; Table 1; Fig. 2). Four zircon fractions from the Cuillin pegmatites yielded concordant or nearly concordant $^{206}\text{Pb}/^{238}\text{U}$ ages ranging from 58.78 ± 0.18 to 58.93 ± 0.14 Myr (Table 1; Fig. 2); the best fractions give a mean $^{206}\text{Pb}/^{238}\text{U}$ age of 58.91 ± 0.07 Myr (2σ ; MSWD = 0.18). The generally high degree of concordancy and tight clustering of the U–Pb data for fractions within both samples argue strongly against significant inheritance of older lead, or recent lead loss. Moreover, the low to moderate U concentration in these zircons (coupled with pre-analysis air abrasion) suggests that lead loss should be negligible. Effects due to $^{230}\text{Th}/^{238}\text{U}$ disequilibrium during zircon growth are also minimal (Table 1).

Phlogopite $^{40}\text{Ar}/^{39}\text{Ar}$ laser step-heating analyses from the Rum pegmatite yield an isochron age of 60.2 ± 1.2 Myr (Fig. 3a), in agreement with the zircon age, suggesting that the zircon age reliably records the cooling of the Rum intrusion past the blocking temperature for Pb diffusion. Thermal degassing of mica crystals from the trachyte dyke intruding the upper part of the Skye lavas yields an

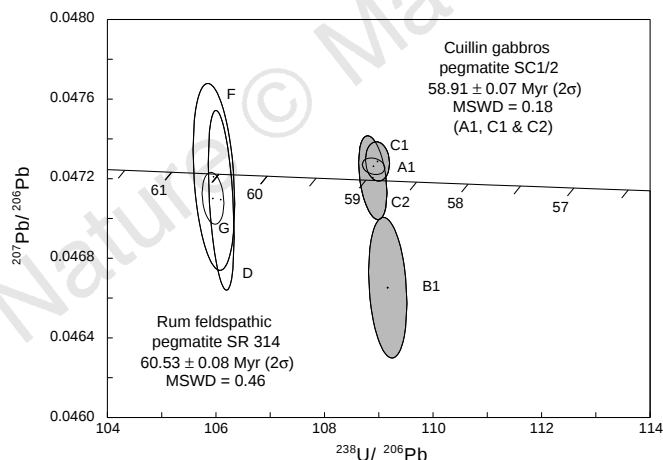


Figure 2 U–Pb concordia diagram showing zircon data from the basic/ultrabasic intrusion complexes of Rum and Cuillin, Skye. Error ellipses are 2σ correlated errors. A weighted mean of all four Cuillin fractions gives a $^{206}\text{Pb}/^{238}\text{U}$ age of 58.89 ± 0.07 Myr (2σ ; MSWD = 0.68). Fraction B1 (<100 pg Pb load) is characterized by minor reverse discordance. Although the $^{206}\text{Pb}/^{238}\text{U}$ age for this fraction is within analytical error of the other ages, it is of comparably poorer quality; exclusion of this fraction results in a preferred mean $^{206}\text{Pb}/^{238}\text{U}$ age for this sample of 58.91 ± 0.07 Myr (2σ ; MSWD = 0.18). Minor normal discordance in fractions A1 and C1 (5.8%, 7.7%) is corrected slightly by consideration of ^{206}Pb deficiency occurring as a result of initial $^{230}\text{Th}/^{238}\text{U}$ disequilibrium during zircon growth (via exclusion of ^{230}Th). Such calculations, using model zircon Th/U of ~ 0.93 – 1.0 , and an assumed parental magma Th/U of ~ 4 , yield negligibly older ages. (These corrections are less than the 2σ internal precisions). Moreover, worse (reverse) discordance is generated for half of the fractions and, taken together with the concordant Rum results, support the contention that the effect here is insignificant.

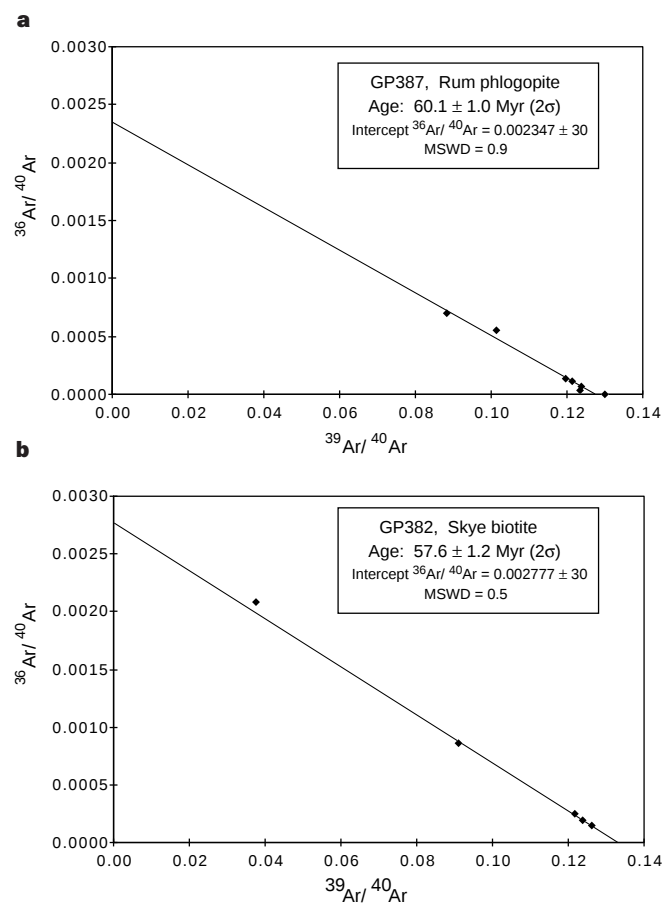


Figure 3 Ar isochron plots of individual laser step-heating gas extractions from Rum harrisite phlogopite [NM3380 9560] (a) and biotite from Skye trachyte dyke [NG 4060 3218] (b). Errors on ages are 2σ . Methods are given in ref. 8. Ages were calculated based on analyses of biotite standard GA1550 assuming an age of 97.9 Myr (ref. 27).

Table 1 U–Pb isotope analyses of zircons

Fraction	Weight (μg)	U (p.p.m.)	Pb* (p.p.m.)	²⁰⁶ Pb/ ²⁰⁴ Pb†	Pb‡ (pg)	²⁰⁸ Pb/ ²⁰⁶ Pb§	²⁰⁶ Pb/ ²³⁸ U§	²⁰⁷ Pb/ ²³⁵ U§	²⁰⁷ Pb/ ²⁰⁶ Pb§	Apparent age (Myr) ²⁰⁶ Pb/ ²³⁸ U
SC-1/2 Pegmatite, Cuillin Gabbro, Skye										
A1	1,848	58	0.63	5,668	11	0.3198	0.009182 ± 0.10	0.05984 ± 0.11	0.04726 ± 0.04	58.92 ± 0.12
B1	131	70	0.75	805	7	0.3154	0.009160 ± 0.16	0.05892 ± 0.44	0.04665 ± 0.38	58.78 ± 0.18
C1	855	58	0.61	1,973	15	0.2978	0.009176 ± 0.15	0.05983 ± 0.15	0.04729 ± 0.10	58.88 ± 0.12
C2	728	63	0.67	621	44	0.3019	0.009184 ± 0.12	0.05977 ± 0.29	0.04721 ± 0.22	58.93 ± 0.14
SR-314 Feldspathic Peridotite, Rum										
D	106	344	4.49	979	22	0.5603	0.009426 ± 0.11	0.06120 ± 0.54	0.04709 ± 0.48	60.48 ± 0.14
F	100	299	3.82	406	46	0.5217	0.009438 ± 0.18	0.06143 ± 0.58	0.04721 ± 0.50	60.56 ± 0.22
G	67	374	5.14	1,287	12	0.6393	0.009439 ± 0.10	0.06130 ± 0.18	0.04710 ± 0.14	60.56 ± 0.12

Errors, 1 standard error of the mean in %. ²⁰⁶Pb/²³⁸U age errors are 2 standard errors in Myr. Pb*, radiogenic Pb. Analytical techniques: Samples abraded to eliminate surface-correlated Pb loss and minimize common Pb corrections. The least-magnetic, optically homogeneous, inclusion- and crack-free grains were analysed. Analytical techniques using a mixed ²⁰⁶Pb/^{233–235}U spike detailed in ref. 28. Total procedural Pb blanks, 4–15 pg (most <6 pg); U <2 pg (most <0.5 pg). Isotopic ratios measured statically on a Finnigan MAT 261 mass spectrometer. Sample descriptions: Rum (SR-314) Zircons from an alkaline segregation 120 m south of Salisbury's dam, within the Rum layered sequence (NM36439970). Generally subhedral grains and fragments between 50 and 200 μm. Good prismatic terminations only rarely noted. Grains were clear, colourless, and commonly showed signs typical of skeletal growth, such as tabular, plate- or flange-like shapes, or having elongate fluid inclusions (parallel to the long axis of the grains). No cores or mineral inclusions were noted in selected grains. Approximately 100 grains analysed for each fraction. Skye (SC1/2) Numerous large (>149 μm), colourless to very pale yellow-brown zircons, recovered from a quartz-feldspathic pegmatite segregation within the Outer Layered Eucrite Series of the Cuillin intrusion, Skye (NG48721963). All grains were of exceptional clarity, comprising either well terminated prisms or fragments of elongate square or flat crystals. Sharply faceted varieties consisted of either equant shapes, or somewhat elongate (length:width ~2–3:1) prisms. Flat grain fragments occasionally showed fluid inclusion trains perpendicular to the longer crystallographic axes. All selected grains were optically free of mineral inclusions or cores. 20–30 grains were chosen for each fraction.

† Measured ratio, corrected for mass fractionation and spike.

‡ Total common Pb in analysis, corrected for fractionation and spike.

§ Corrected for spike, blank and common Pb.

⁴⁰Ar–³⁹Ar isochron age of 57.6 ± 1.2 Myr (Fig. 3b), close to the Cuillin zircon age.

Field and geochronological information can be used to estimate magma production rates. The oldest, most reliably dated lavas in the British Tertiary Volcanic Province are those from the Eigg Lava Formation, exposed on Muck. Sanidines from a tuff horizon near the base of these lavas yield ⁴⁰Ar–³⁹Ar plateau ages⁸ of 62.4 ± 0.6 to 62.8 ± 0.6 Myr. Field relations show that these tuffs and lavas clearly pre-date the Rum ultrabasic intrusion (Fig. 1a). This is confirmed by the new age determinations (Table 1; Fig. 1b) which firmly establish that eruption of the SMLS began during unroofing and erosion of the Rum central intrusive complex, following its formation at 60.5 Myr (Fig. 1). The Cuillin complex intrudes the SMLS and fed the Preshal-More dykes and lavas^{15,16}. Therefore the 58.9 Myr Cuillin U–Pb age provides an upper age bound for the lavas, suggesting a duration of 1.6 ± 0.2 Myr for formation of the lava pile. This timespan is similar to that suggested for the eruption of the Mull lavas, based on ⁴⁰Ar–³⁹Ar data¹⁷, although the errors of the latter are much larger (1.5 ± 4.6 Myr). Our precisely resolved maximum eruption interval is comparable to that observed for numerous other flood basalt provinces (see, for example, refs 4 and 18). The exposed Skye lava field is estimated to be between 1.2 and 1.5 km thick^{6,7} and is unlikely to be significantly thicker beneath the Canna basin⁶. A maximum timespan of 1.6 Myr for the main Skye volcanic centre suggests magma production volumes of the order of 2.2 × 10^{–3} km³ yr^{–1}, assuming 1,400 km³ for the volume of the SMLS and accounting for basic/ultrabasic rocks thought to have formed the Cuillin centre and its roots³. The eruption interval of 0.2 Myr proposed by Jolley¹⁹ gives an order-of-magnitude greater eruption rate. If significant volumes of magma were underplated at the base of the crust (3–5 times extruded volumes)⁴, these rates would be significantly greater. These magma production rates are high and compatible with those associated with individual magmatic centres above an upwelling plume, such as that beneath Hawaii⁴.

Depths of origin of basic magmas can be estimated using trace-element inversion models and phase-equilibria data. The SMLS is transitional to alkalic in chemistry¹⁰ and, when the effects of crustal contamination are removed, the lava compositions are similar to ocean island basalt (OIB) magmas. Trace-element characteristics, in particular the relative depletion of heavy rare-earth elements, indicate melt segregation beneath relatively thick (70–80 km) lithosphere^{4,13}. The pressure-dependent melting of anhydrous mantle peridotite is well constrained by experiments²⁰ and application of the data to the SMLS independently support the conclusions from trace-element modelling^{21,22}; that is, most magma segregated

from ~80 km depth. Petrological studies indicate that the parental magma to the Rum intrusion was alkalic in nature²³, further substantiating a relatively thick lithosphere beneath the region in pre-, MLS times. The Preshal-More magma type is tholeiitic, and exhibits strong chemical similarity to MORB². It is widely accepted that the formation of MORB-like magmas requires larger degrees of partial melting and a relatively thin lithosphere, with most melt segregating from <50 km (refs 4, 20). Thus, the change in magma composition being erupted on Skye, from OIB-like to MORB-like, suggests thinning of continental lithosphere to around 50 km (refs 4, 13). A similar history of lithospheric thinning, based on less precise chronology (Fig. 1b), has been inferred on the basis of trace-element geochemistry for the area beneath Mull, 60 km south of Skye²⁴.

The geochronology presented here places stringent time constraints on the rate of proposed lithospheric thinning beneath the Skye area. The lava chemistry implies a loss of at least 20 km of lithosphere over a period of around 1.6 Myr. Although the timescale of this process is similar to that inferred for Mull, it is not yet clear whether such thinning took place only locally at the magmatic centres or over the entire region, between Skye and Mull. The most precise age estimate¹⁷ for the Mull plateau lavas, 60.0 ± 1.0 Myr, is coincident with the period that we suggest for the eruption of Skye lavas (Fig. 1b; but we note large uncertainty and a possible younger biostratigraphic age of 55 Myr, ref. 19). To this extent, the lithospheric thinning appears synchronous beneath these areas. These observations then prompt the question of whether the implied magnitude of lithospheric thinning can be driven purely by extension, or whether it reflects additional erosion by a plume head²⁴. We also note that the high-precision U–Pb ages presented here may be used with existing magnetostratigraphy to provide valuable tie-points for the palaeomagnetic timescale (Fig. 1b). □

Received 2 January; accepted 6 April 1998.

1. Harker, A. *The Tertiary Igneous Rocks of Skye* (HMSO, Edinburgh, 1904).
2. Thompson, R. N. Magmatism in the British Tertiary Volcanic Province. *Scott. J. Geol.* **18**, 49–107 (1982).
3. Emeleus, C. H. & Gyopari, M. C. *British Tertiary Volcanic Province* (Chapman & Hall, London, 1992).
4. White, R. S. & McKenzie, D. P. Mantle plumes and flood basalts. *J. Geophys. Res.* **100**, 17543–17585 (1995).
5. Anderson, F. W. & Dunham, K. C. *The Geology of Northern Skye* (HMSO, Edinburgh, 1966).
6. England, R. W. The structure of the Skye Lava field. *Scott. J. Geol.* **30**, 33–37 (1994).
7. Williamson, I. T. & Bell, B. R. The Palaeocene lava field of west-central Skye, Scotland: stratigraphy, palaeogeography and structure. *Trans. R. Soc. Edinb.: Earth Sci.* **85**, 39–75 (1994).
8. Pearson, D. G., Emeleus, C. H. & Kelley, S. P. Precise ⁴⁰Ar/³⁹Ar age for the initiation of Palaeogene volcanism in the Inner Hebrides and its regional significance. *J. Geol. Soc. Lond.* **158**, 815–818 (1996).
9. Saunders, A. D., Fitton, J. G., Kerr, A. C., Norry, M. J. & Kent, R. W. In *Large Igneous Provinces* (eds Mahoney, J. & Coffin, M. F.) 45–93 (Geophys. Monogr. 100, Am. Geophys. Union, Washington DC, 1997).
10. Thompson, R. N., Esson, J. & Dunham, A. C. Major element chemical variation in the Eocene lavas of the Isle of Skye, Scotland. *J. Petrol.* **13**, 219–253 (1972).

11. Matthey, D. P., Gibson, I. L., Marriner, G. F. & Thompson, R. N. The diagnostic geochemistry, relative abundance and spatial distribution of high-calcium, low-alkali olivine tholeiite dykes in the Lower Tertiary regional dyke swarm of the Isle of Skye, NW Scotland. *Mineral. Mag.* **41**, 273–285 (1977).
12. Thompson, R. N., Gibson, I. L., Marriner, G. F., Matthey, D. P. & Morrison, M. A. Trace element evidence of multi-stage mantle fusion and polybaric fractional crystallisation in the Palaeocene lavas of Skye, NW Scotland. *J. Petrol.* **21**, 265–293 (1980).
13. Ellam, R. M. Lithospheric thickness as a control on basalt geochemistry. *Geology* **20**, 153–156 (1992).
14. Meighan, I., Hutchison, R., Williamson, I. & Macintyre, R. M. Geological evidence for the different relative ages of the Rhum and Skye Tertiary central complexes. (abstr.) *J. Geol. Soc. Lond.* **139**, 659 (1981).
15. Dickin, A. P., Brown, J. L., Thompson, R. N., Halliday, A. N. & Morrison, M. A. Crustal contamination and the granite problem in the British Tertiary Volcanic Province. *Phil. Trans. R. Soc. Lond. A* **310**, 755–780 (1984).
16. Walker, G. P. L. in *Magmatic Processes and Plate Tectonics* (eds Pritchard, H. M., Alabaster, T., Harris, N. B. W. & Neary, C. R.) 489–497 (Spec. Publ. 76, Geol. Soc., London, 1993).
17. Musset, A. E. ⁴⁰Ar–³⁹Ar step heating ages of the Tertiary igneous rocks of Mull, Scotland. *J. Geol. Soc. Lond.* **143**, 887–896 (1986).
18. Renne, P. R., Deckart, K., Ernesto, M., Feraud, G. & Piccirillo, E. M. Age of the Ponta-Grossa dike swarm (Brazil), and implications to Parana flood volcanism. *Earth Planet. Sci. Lett.* **144**, 199–211 (1996).
19. Jolley, D. W. in *Palaeosurfaces: Recognition, Reconstruction & Palaeoenvironmental Interpretation* (ed. Widdowson, M.) 67–94 (Spec. Publ. 120, Geol. Soc., London, 1997).
20. Langmuir, C. H., Klein, E. M. & Plank, T. in *Mantle Flow and Melt Generation at Mid-Ocean Ridges* (eds Phipps-Morgan, J., Blackman, D. K. & Sinton, J. M.) 183–281 (Geophys. Monogr. 71, Am. Geophys. Union, Washington DC, 1992).
21. Thompson, R. N. Primary basalts and magma genesis. *Contrib. Mineral. Petrol.* **45**, 317–341 (1974).
22. Searrow, J. H. & Cox, K. G. Basalts generated by decompressive adiabatic melting of a mantle plume: a case study from the Isle of Skye, NW Scotland. *J. Petrol.* **36**, 3–22 (1995).
23. Kitchen, D. E. The parental magma on Rhum: evidence from alkaline segregations and veins in the peridotite from Salisbury's Dam. *Geol. Mag.* **122**, 529–537 (1985).
24. Kerr, A. C. Lithospheric thinning during the evolution of continental large igneous provinces: A case study from the North Atlantic province. *Geology* **22**, 1027–1030 (1994).
25. Dickin, A. P. Isotope geochemistry of Tertiary igneous rocks from the Isle of Skye. *J. Petrol.* **22**, 155–189 (1981).
26. Berggren, W., Kent, D. V., Swisher, C. C. & Aubury, M.-P. in *Geochronology, Time Scales and Global Stratigraphic Correlation* (eds Berggren, W., Kent, D. V., Aubury, M.-P. & Hardenbol, J.) 129–212 (Spec. Publ., Soc. Economic Palaeontologists and Mineralogists, 1995).
27. McDougall, I. & Harrison, T. M. *Geochronology and Thermochronology by the ⁴⁰Ar/³⁹Ar Method* (Oxford Univ. Press, 1988).
28. Parrish, R. R., Roddick, J. C., Loveridge, W. D. & Sullivan, R. W. Uranium-lead analytical techniques at the geochronology laboratory. *Geol. Surv. Can. Pap.* **87-2**, 3–7 (1987).

Acknowledgements. We thank J. Wartho for assistance with the laser argon extractions, and the staff of the GSC's Geochronology laboratory for help with the U-Pb analyses. Reviews from P. Renne, R. Duncan and K. Ludwig and S. Bergman improved the manuscript.

Correspondence and requests for materials should be addressed to D.G.P. (e-mail: d.g.pearson@durham.ac.uk).

Influence of mesoscale eddies on new production in the Sargasso Sea

D. J. McGillicuddy Jr^{*}, A. R. Robinson[†], D. A. Siegel[‡], H. W. Jannasch[§], R. Johnson^{||}, T. D. Dickey[¶], J. McNeill[¶], A. F. Michaels[#] & A. H. Knap^{||}

^{*} Department of Applied Ocean Physics and Engineering, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA

[†] Department of Earth and Planetary Sciences, Harvard University, Cambridge, Massachusetts 02138, USA

[‡] Institute for Computational Earth System Science and Department of Geography, [§] Ocean Physics Laboratory, University of California, Santa Barbara, California 93106, USA

[¶] Monterey Bay Aquarium Research Institute, Moss Landing, California 95039, USA

^{||} Bermuda Biological Station for Research, Ferry Reach, GE01, Bermuda

[#] Wrigley Institute for Environmental Studies, University of Southern California, Avalon, California 90704, USA

It is problematic that geochemical estimates of new production—that fraction of total primary production in surface waters fuelled by externally supplied nutrients—in oligotrophic waters of the open ocean surpass that which can be sustained by the traditionally accepted mechanisms of nutrient supply.^{1,2} In the case of the Sargasso Sea, for example, these mechanisms account for less than half of the annual nutrient requirement indicated by new production estimates based on three independent transient-tracer tech-

niques^{2–6}. Specifically, approximately one-quarter to one-third of the annual nutrient requirement can be supplied by entrainment into the mixed layer during wintertime convection⁷, with minor contributions from mixing in the thermocline^{8,9} and wind-driven transport¹⁰ (the potentially important role of nitrogen fixation¹¹—for which estimates vary by an order of magnitude in this region¹²—is excluded from this budget). Here we present four lines of evidence—eddy-resolving model simulations, high-resolution observations from moored instrumentation, ship-board surveys and satellite data—which suggest that the vertical flux of nutrients induced by the dynamics of mesoscale eddies is sufficient to balance the nutrient budget in the Sargasso Sea.

The notion that mesoscale (of the order of 10² km) processes could be an important vehicle for nutrient transport has been debated for some time^{13,14}. Analysis of an apparently eddy-driven event observed at Station S near Bermuda in 1986 suggested that only a few such events per year would be required to account for the total new production¹⁵. A variety of mechanisms have been proposed as causes of vertical nutrient flux, from the spindown of anticyclonic vortices¹⁶ to submesoscale upwelling patches in meandering jets^{13,17}. Modelling studies in the northeast Atlantic Ocean^{18,19} suggested substantial nutrient fluxes associated with cyclonic eddies and their interactions with adjacent features. Recent work²⁰ has indicated that mesoscale eddies could be the dominant mode of nutrient transport in the open ocean.

Using a regional eddy-resolving model, McGillicuddy and Robinson²⁰ constructed long-term seasonal simulations characteristic of the mesoscale environment in the Sargasso Sea. Detailed comparisons with observations (eddy kinetic energies, space and time scales) demonstrate that the simulations contain statistically realistic representations of mesoscale fluctuations. Incorporation of a simplified biological model into these calculations facilitated study of the mechanisms by which eddy processes can transport nutrients into the euphotic zone. Results show²⁰ that upwelling due to the formation of cyclonic eddies and subsequent intensification caused by interaction with surrounding features cause sporadic nutrient injections into the surface layer. These calculations indicate the annual flux resulting from the eddy upwelling process in the

Table 1 Geochemical estimates of new production in the Sargasso Sea and nitrate sources

Method		Value (mol N m ⁻² yr ⁻¹)	Reference
O ₂ utilization		0.48 ± 0.10	Ref. 5
		0.42 ± 0.09	Ref. 2
	O ₂ production	0.46 ± 0.09	Ref. 2
		0.39 ± 0.16	Ref. 3
³ He flux gauge		0.51 ± 0.14	Ref. 3
		0.56 ± 0.16	Ref. 6
		0.47 ± 0.15	W. J. Jenkins, personal communication
		0.70 ± 0.20	Ref. 27
Nitrogen demand		0.50 ± 0.14	
Nitrate sources			
Process	Method	Value (mol N m ⁻² yr ⁻¹)	Reference
Wintertime convection	O ₂ production	0.17 ± 0.05	Ref. 7
	NO ₃ removal	0.09 ± 0.04	Ref. 7
Diapycnal diffusion	Microstructure	0.05 ± 0.01	Ref. 8
Ekman flow	Climatological hydrography and winds	0.03 ± 0.01	Ref. 10
Eddy upwelling	Simulation	0.35 ± 0.10	Ref. 20
	Satellite-based statistical model	0.19 ± 0.10	This work
	Nitrate supply	0.48 ± 0.17*	

* Excludes nitrogen fixation, for which estimates are very poorly constrained (ranging from 0.05 to 1.3 mol N m⁻² yr⁻¹ in the Sargasso Sea²⁸).

Sargasso Sea is $0.35 \pm 0.1 \text{ mol N m}^{-2} \text{ yr}^{-1}$, which is sufficient to reconcile the apparent discrepancy in the nutrient budget described above (Table 1).

We consider the following conceptual model of the eddy upwelling mechanism (Fig. 1). A density surface of mean depth Z_0 is coincident with the depth of the euphotic zone. This density surface is perturbed by the formation, evolution and destruction of meso-scale features. Nutrients injected into the euphotic zone by shoaling density surfaces are fixed by the biota, whereas deepening density surfaces serve to push nutrient-depleted water out of the well-lit surface layers. The asymmetry imposed by the light field thus rectifies vertical displacements (both up and down) into a net upward transport of nutrients.

Recent advances in the theory of advective effects on planktonic ecosystems provide a context in which to examine analytically the biological response to injection events²¹. Arrival of nutrients to a particular depth in the light-saturated layer stimulates phytoplankton growth which is at first linear. Nutrients begin to accumulate and then decrease rapidly as the phytoplankton population enters a phase of exponential growth. Using parameters relevant to the Sargasso Sea, the response time is one to several days. These findings are in line with earlier work demonstrating biological capability for rapid utilization of episodic nutrient inputs²², and are consistent with the fact that near-surface nitrate concentrations in this region remain below the limit of detection except during periods of wintertime convection.

Given that the eddy driven nutrient flux is such a large component of the annual budget, it is surprising that this mechanism could function largely undetected by biweekly to monthly discrete water sampling in shipboard time-series operations off Bermuda. Monthly observations tend to undersample this highly sporadic process²⁰. Furthermore, the time required for biological removal of new nutrient is much quicker (days) than that of the supply mechanism (weeks). This dichotomy in timescales makes it very difficult to observe evidence of a nutrient injection directly with traditional shipboard hydrographic methods.

High-resolution time series using moored instrumentation provides an observational approach capable of resolving these intermittent events. The Bermuda Testbed Mooring has been deployed since June 1994 ~80 km southeast of that island, near the Bermuda Atlantic Time Series (BATS) site²³. During the summer of 1995, an eddy event was observed (Fig. 2). Dramatic cooling began on day 185 of the year and persisted for 30 days. The record from an automated nitrate analyser²⁴ placed near the base of the euphotic

zone showed nutrient enhancement associated with this event, in which nitrate concentrations rose from undetectable to 1.4 mmol m^{-3} . This was accompanied by increases in both chlorophyll fluorescence and beam attenuation coefficient, indicating high concentrations of both phytoplankton biomass and particulate material. Chlorophyll during this period was apparently the highest observed in the BATS program to date.

The magnitude and duration of these temporal changes in physical and biogeochemical properties is consistent with the eddy upwelling mechanism described above. The eulerian timescale for the passage of such features can be calculated by dividing a typical eddy diameter (~150 km) by a characteristic propagation speed ($\sim 5 \text{ km d}^{-1}$; ref. 25) which yields an estimate of one month (precisely that of the observed event). The magnitude of the temperature decrease and associated nitrate enhancement at the base of the euphotic zone are consistent with an isopycnal displacement of ~80 m. Assuming complete nitrate utilization, only four events of this size are needed each year to provide the annual budget.

Monthly sampling during time-series operations at the BATS site is also supplemented by periodic 'validation' cruises which are used to provide spatial context for the fixed-point observations. One such survey was conducted in June 1996 (Fig. 3). The hydrographic structure included two cold features in the northwestern and eastern portions of the domain, and a warm anomaly to the south. These structures had a dramatic influence on the nitrate distribution just below the euphotic zone. Nitrate concentrations were below the limit of detection in the core of the warm feature, and in excess of 2 mmol m^{-3} in the interiors of the two cold anomalies. Phytoplankton biomass patterns (indicated by chlorophyll) within the euphotic zone corresponded to the underlying nitrate distribution. Although

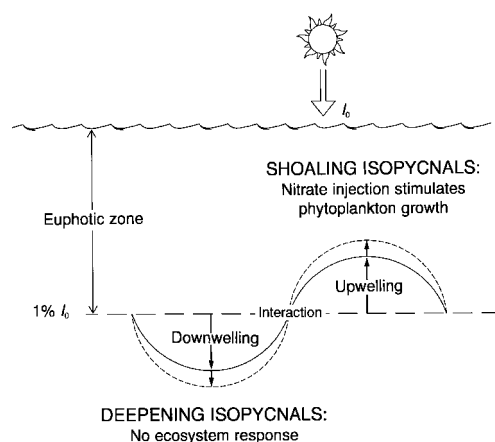


Figure 1 A schematic representation of the eddy upwelling mechanism. The solid line depicts the vertical deflection of an individual isopycnal caused by the presence of two adjacent eddies of opposite sign. The dashed line indicates how the isopycnal might be subsequently perturbed by interaction of the two eddies. I_0 represents incident solar radiation, and $1\% I_0$ the base of the euphotic zone.

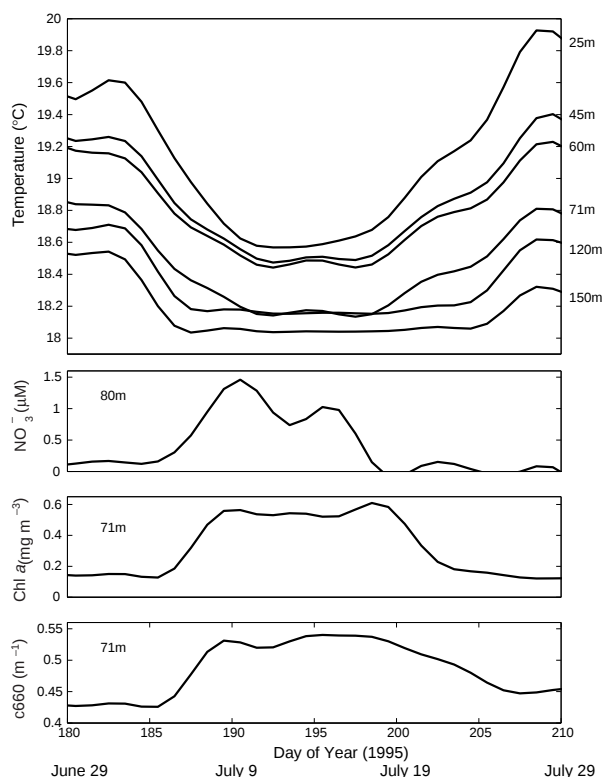


Figure 2 Results from Bermuda Testbed Mooring deployment 3 during the summer of 1995. Upper panel, temperature records at various depths; lower three panels, nitrate concentration at 80 m, chlorophyll fluorescence and beam attenuation coefficient (c_{660}) at 71 m. All signals have been filtered via a six-day moving average. Chlorophyll values of the broadband data exceed 1.4 mg m^{-3} .

the correlation is not exact, biomass is generally lower above the warmer nutrient-depleted waters, and higher above colder more nutrient-rich areas.

Given the conceptual model of the eddy upwelling mechanism, it is possible to estimate nutrient fluxes based on altimetric determinations of the eddy field. Dynamical and empirical mode analyses demonstrate that the vertical structure of these eddy signals are dominated by the first baroclinic mode²⁵, enabling the use of satellite-derived sea-level anomaly (SLA) data to infer sub-surface isopycnal displacements. At the BATS site, isopycnal displacements at 300 m depth are correlated with Topex/Poseidon SLA ($r^2 = 0.62$) with a slope of 4 m of isopycnal displacement per 1 cm of SLA. Using

this information and the mean relationship between density and nitrate concentration derived from BATS data, the upward flux of nitrate across the base of the euphotic zone is estimated to be $0.19 \pm 0.1 \text{ mol N m}^{-2} \text{ yr}^{-1}$ (Table 1). This estimate is roughly consistent with, although smaller than, the results obtained by McGillicuddy and Robinson²⁰.

Although this empirically derived relationship between SLA and isopycnal displacement at the base of the euphotic zone is statistically significant, this simple model cannot accurately represent the complex vertical structure which is sometimes observed. In fact, analysis of the hydrographic time series bracketing the mooring record described above reveals that the doming of the seasonal thermocline (which resulted in the nitrate pulse and exceptionally high chlorophyll) was accompanied by a depression of the main thermocline. The net effect results in a positive SLA, as opposed to the negative signature characteristic of cyclones. Thus, nitrate flux associated with features of this type is not accounted for in this approach. However, similar vertical structures have been observed only four times in the nine-year BATS hydrographic record (although sampling remains an issue). Nevertheless, this statistical model should underestimate new production. Further synthesis of *in situ* observations with remote sensing in the context of data-assimilative dynamical models offers an opportunity to better understand the role of intermittent processes in biogeochemical cycling in the ocean.

We point out that the mechanism presented here cannot account for the summertime drawdown of dissolved inorganic carbon observed at BATS. Eddy induced upwelling of nitrate will be accompanied by a commensurate flux of dissolved inorganic carbon in Redfield proportion. The findings reported here therefore tend to exacerbate previously described problems in balancing carbon budgets in the Sargasso Sea²⁶. □

Received 5 November 1997; accepted 30 April 1998.

- Schulenberg, E. & Reid, J. L. The Pacific shallow oxygen maximum, deep chlorophyll maximum, and primary productivity, reconsidered. *Deep-Sea Res.* **A** *28*, 901–919 (1981).
- Jenkins, W. J. & Goldman, J. C. Seasonal oxygen cycling and primary production in the Sargasso Sea. *J. Mar. Res.* **43**, 465–491 (1985).
- Spitzer, W. S. & Jenkins, W. J. Rates of vertical mixing, gas exchange and new production: estimates from seasonal gas cycles in the upper ocean near Bermuda. *J. Mar. Res.* **47**, 169–196 (1989).
- Sarmiento, J. L., Thiele, G., Key, R. M. & Moore, W. S. Oxygen and nitrate new production and remineralization in the North Atlantic subtropical gyre. *J. Geophys. Res.* **95**, 18303–18315 (1993).
- Jenkins, W. J. & Wallace, D. W. R. in *Primary Productivity and Biogeochemical Cycles in the Sea* (eds Falkowski, P. G. & Woodhead, A. D.) 299–316 (Plenum, New York, 1992).
- Jenkins, W. J. Nitrate flux into the euphotic zone near Bermuda. *Nature* **331**, 521–523 (1988).
- Michaels, A. F. et al. Seasonal patterns of ocean biogeochemistry at the U.S. JGOFS Bermuda Atlantic Time-series Study site. *Deep-Sea Res.* **141**, 1013–1038 (1994).
- Lewis, M. R., Harrison, W. G., Oakley, N. S., Hebert, D. & Platt, T. Vertical nitrate fluxes in the oligotrophic ocean. *Science* **234**, 870–873 (1986).
- Ledwell, J. R., Watson, A. J. & Law, C. S. Evidence for slow mixing across the pycnocline from an open-ocean tracer-release experiment. *Nature* **364**, 701–703 (1993).
- Williams, R. G. & Follows, M. J. The Ekman transfer of nutrients and maintenance of new production over the North Atlantic. *Deep-Sea Res.* **145**, 461–489 (1998).
- Karl, D. et al. The role of nitrogen fixation in biogeochemical cycling in the subtropical North Pacific Ocean. *Nature* **388**, 533–538 (1997).
- Michaels, A. F. et al. Inputs, losses and transformations of nitrogen and phosphorus in the pelagic North Atlantic Ocean. *Biogeochemistry* **35**, 181–226 (1996).
- Woods, J. D. in *Toward a Theory on Biological-Physical Interactions in the World Ocean* (ed. Rothschild, B. J.) (Reidel, Dordrecht, 1988).
- Falkowski, P. G., Ziemann, D., Kolber, Z. & Bienfang, P. K. Role of eddy pumping in enhancing primary production in the ocean. *Nature* **353**, 55–58 (1991).
- Jenkins, W. J. The use of anthropogenic tritium and helium-3 to study subtropical gyre ventilation and circulation. *Phil. Trans. R. Soc. Lond. A* **325**, 43–61 (1988).
- Franks, P. J. S., Wroblewski, J. S. & Flierl, G. R. Prediction of phytoplankton growth in response to the frictional decay of a warm-core ring. *J. Geophys. Res.* **91**, 7603–7610 (1986).
- Strass, V. H. Chlorophyll patchiness caused by mesoscale upwelling at fronts. *Deep-Sea Res.* **39**, 75–96 (1992).
- McGillicuddy, D. J., Robinson, A. R. & McCarthy, J. J. Coupled physical and biological modeling of the spring bloom in the North Atlantic: (ii) three dimensional bloom and post-bloom effects. *Deep-Sea Res.* **142**, 1359–1398 (1995).
- Robinson, A. R. et al. Mesoscale and upper ocean variabilities during the 1989 JGOFS bloom study. *Deep-Sea Res.* **140**, 9–35 (1993).
- McGillicuddy, D. J. & Robinson, A. R. Eddy induced nutrient supply and new production in the Sargasso Sea. *Deep-Sea Res.* **144**, 1427–1449 (1997).
- Robinson, A. R. On the theory of advective effects on biological dynamics in the sea. *Proc. R. Soc. Lond. A* **453**, 1–30 (1997).
- Goldman, J. C. Potential role of large oceanic diatoms in new primary production. *Deep-Sea Res.* **140**, 159–168 (1993).
- Dickey, T. et al. Initial results from the Bermuda Testbed Mooring program. *Deep-Sea Res.* **145**, 771–794 (1998).

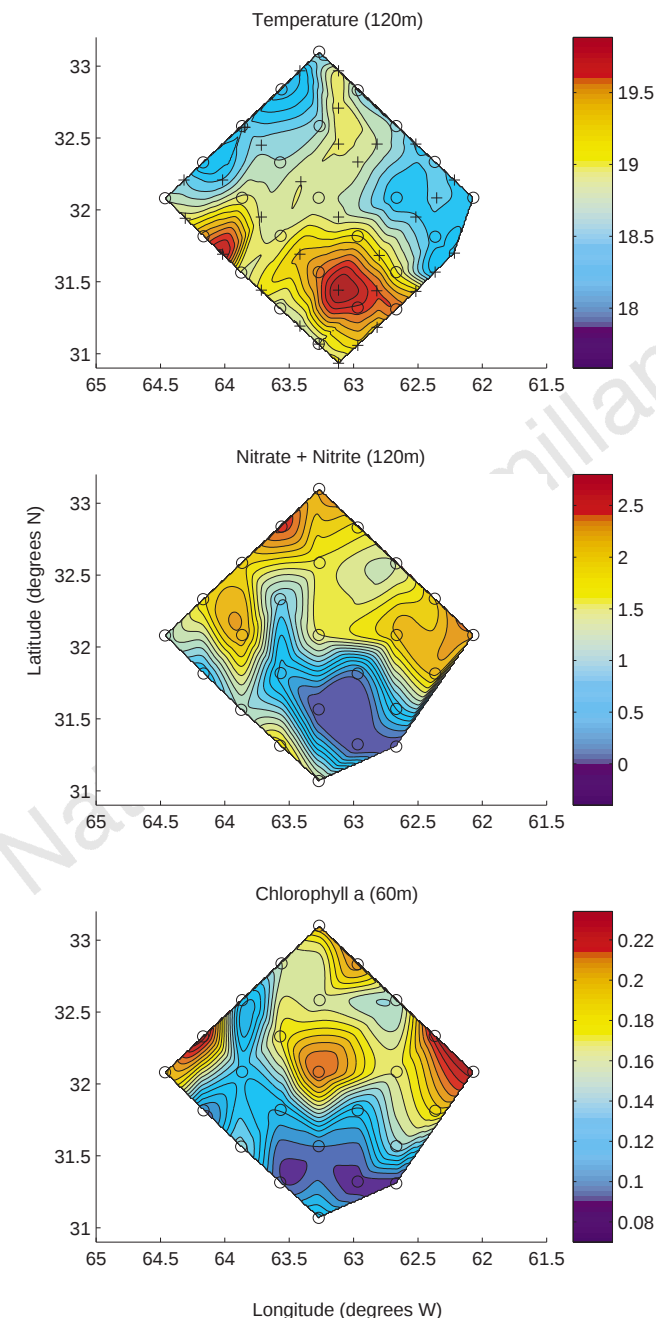


Figure 3 Results from a mesoscale biogeochemical survey near the BATS site occupied from 24 to 28 June 1996. Top, temperature at 120 m (°C); middle, nitrate + nitrite at 120 m (mmol m⁻³); bottom, chlorophyll *a* at 60 m (mg m⁻³). Circles and crosses indicate hydrographic stations and expendable bathythermograph casts, respectively.

24. Jannasch, H. W., Johnson, K. S. & Sakamoto, C. M. Submersible osmotically pumped analyzers for continuous determination of nitrate *in situ*. *Anal. Chem.* **66**, 3352–3361 (1994).
25. Richman, J. G., Wunsch, C. & Hogg, N. G. Space and time scales of mesoscale motion in the Western North Atlantic. *Rev. Geophys.* **15**, 385–420 (1977).
26. Michaels, A. F., Bates, N. R., Buesseler, K. O., Carlson, C. A. & Knap, A. H. Carbon cycle imbalances in the Sargasso Sea. *Nature* **372**, 537–540 (1994).
27. Jenkins, W. J. Studying subtropical thermocline ventilation and circulation using tritium and ^3He . *J. Geophys. Res.* (submitted).
28. Lipschultz, F. & Owens, N. J. P. An assessment of nitrogen fixation as a source of nitrogen for the North Atlantic Ocean. *Biogeochemistry* **35**, 261–274 (1996).

Acknowledgements. We thank W. Jenkins, J. Goldman, S. Emerson, P. Cornillon and J. Yoder for discussions; E. Fields for his contribution to the processing and analysis of the various data sets; and the BATs technicians for their assistance. This work was supported by JPL, NASA, the US NSF, NOAA and ONR.

Correspondence and requests for materials should be addressed to D.J.McG. (e-mail dmccillicuddy@whoi.edu).

Eddy-induced enhancement of primary production in a model of the North Atlantic Ocean

Andreas Oschlies*† & Véronique Garçon*

* UMR5566/LEGOS, Centre National de la Recherche Scientifique, 18 Avenue Edouard Belin, 31401 Toulouse Cedex 4, France

† Institut für Meereskunde an der Universität Kiel, Düsterbrook Weg 20, 24105 Kiel, Germany

In steady state, the export of photosynthetically fixed organic matter to the deep ocean has to be balanced by an upward flux of nutrients into the euphotic zone¹. Indirect geochemical estimates² of the nutrient supply to surface waters have been substantially higher than direct biological and physical measurements³, particularly in subtropical regions. A possible explanation for the apparent discrepancy is that the sampling strategy of the direct measurements has under-represented episodic nutrient injections forced by mesoscale eddy dynamics, whereas geochemical tracer budgets integrate fluxes over longer time and space scales. Here we investigate the eddy-induced nutrient supply by combining two methods potentially capable of delivering synoptic descriptions of the ocean's state on a basin scale. Remotely sensed sea-surface height data from the simultaneous TOPEX/Poseidon and ERS-1 satellite missions are assimilated into a numerical eddy-resolving coupled ecosystem–circulation model of the North Atlantic Ocean. Our results indicate that mesoscale eddy activity accounts for about one-third of the total flux of nitrate into the euphotic zone (taken to represent new production) in the subtropics and at mid-latitudes. This contribution is not sufficient to maintain the observed primary production in parts of the subtropical gyre, where alternative routes of nitrogen supply will have to be considered.

Assuming that nitrogen is the nutrient limiting biological production in the well-lit upper ocean, all primary production associated with the newly available nitrogen is called new production⁴. Geochemical estimates of space- and time-averaged new production, based on oxygen production and consumption or using tritiogenic ^3He as a flux gauge², have been significantly higher than local and instantaneous measurements of nitrate uptake in incubation experiments, which in turn were consistent with turbulent flux estimates based on microstructure measurements³. Intermittent nutrient pulses by eddy-induced upwelling have been suggested to resolve the observational discrepancy. There is evidence both from observations⁵ and idealized model studies^{6,7} that cyclonic flow anomalies, by raising isopycnal surfaces, can lead to local upwelling of nutrient-rich water into the euphotic zone. Eddies might also enhance the lateral transfer of nutrients from the nutrient-rich regions surrounding the subtropical gyre⁸.

To investigate the influence of mesoscale eddies on biological

production in a realistic environment on a basin scale, a four-component pelagic ecosystem model is coupled to a $(1/3)^\circ$ resolution model of the North Atlantic circulation⁹ derived from the WOCE Community Modeling Effort¹⁰. The numerical model has 37 levels in the vertical, with a grid spacing increasing from 11 m at the surface to 250 m at depth. It is forced with monthly climatological data sets, and lateral boundaries are closed with buffer zones at which temperature and salinity are relaxed to observed climatological values¹¹. In order to better represent the mixed-layer dynamics, a turbulent kinetic energy closure scheme has been implemented¹². Optical properties of the water column are modelled by an analytical formula¹³ for clear ocean water plus a simple exponential absorption law taking into account phytoplankton self-shading.

The model framework for the plankton dynamics is a classical N-P-Z-D (nitrate, phytoplankton, zooplankton, detritus) nitrogen-based biological model. Processes within the trophic chain include phytoplankton growth and mortality, zooplankton growth by grazing

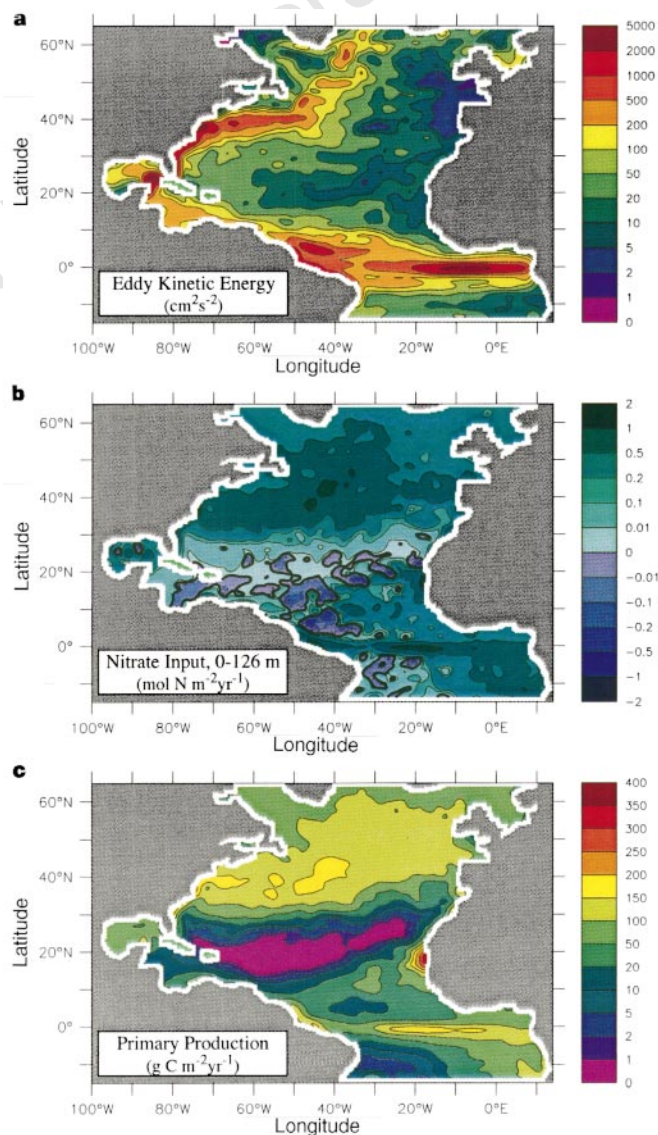


Figure 1 Results from the assimilation experiment A. **a**, Surface eddy kinetic energy (EKE), which contains all deviations from the annual mean, computed for a depth of 60 m to avoid contamination by shallow Ekman currents (in $\text{cm}^2 \text{s}^{-2}$). **b**, Annual mean nitrate flux into the upper 126 m, which is taken as proxy for the euphotic zone (in $\text{mol N m}^{-2} \text{yr}^{-1}$). **c**, Annual mean primary production (in $\text{g C m}^{-2} \text{yr}^{-1}$). A constant ratio of C:N=6.6 was assumed to give carbon fluxes from the model. This is a rather conservative assumption²⁷ and will give minimal estimates of carbon fluxes.

on phytoplankton, excretion and mortality, and remineralization and sinking of detrital material, with uniform biological parameters for the entire basin⁹. The biology is inserted after a 25-year spin-up of the physical model. Initial nitrate concentrations are taken from climatological data¹⁴. Advection of the ecosystem components is modelled with a higher-order positive definite scheme¹⁵, rather than with the classical upstream differencing with its inherent large implicit diffusivity. Although a stable pattern of the seasonal cycle is established after the first year of the coupled integration, the model analysis phase is chosen to be the third coupled year.

First, an eddy-permitting simulation C (control) was run using biharmonic horizontal friction and diffusion (with friction coefficient $A_h = 2.5 \times 10^{19} \text{ cm}^4 \text{ s}^{-1}$) that restricts to small scales the dissipation necessary only for numerical stability reasons. Despite its fine grid resolution, the model still underestimates mesoscale variability over large regions of the North Atlantic¹⁶. To achieve a more realistic representation of the real ocean's eddy field, we set up experiment A (assimilation): mapped sea-surface height observations taken at 5-day intervals from the combined TOPEX/Poseidon and ERS-1 altimetric missions¹⁷, and covering the North Atlantic basin between 8° and 65°N, were assimilated into the numerical model over a one-year period beginning in October 1992. The assimilation method¹⁸ corrects the otherwise unperturbed model evolution every five days by computing a full three-dimensional adjustment from the observed surface model-data misfits. This correction is virtually in geostrophic balance, and does not change temperature or salinity on isopycnals. It further conserves the total amount of nitrogen for each vertical profile of any biological

variable. In sensitivity experiment V (viscous) we effectively reduced the eddy activity in the model by switching to harmonic viscosity (but retaining biharmonic diffusion) with a large friction coefficient ($K_h = 10^8 \text{ cm}^2 \text{ s}^{-1}$) that corresponds to values typically used in coarse-resolution models^{11,19}. On a scale of 50 km, friction in run V is already a hundred times stronger than in experiments A and C.

The surface eddy kinetic energy (EKE) of the assimilation experiment A (Fig. 1a) shows high values associated with the Gulf Stream and its continuation as North Atlantic Current, the Loop Current in the Gulf of Mexico, and the equatorial current system with its strong seasonal variations. Overall, the regional pattern of EKE agrees well with the satellite observations¹⁸. The numerical grid is, however, not yet fine enough to resolve the entire spectrum of mesoscale variability. Even in the assimilation experiment A, simulated EKE levels are systematically too low, typically by a factor of two (Fig. 2a, Table 1).

Although the model does not explicitly distinguish between new and regenerated nitrogen, the net annual flux of nitrate into the euphotic zone (Fig. 1b) will serve as a proxy for new production⁴. Some of the small-scale structure is due to the short integration period determined by the single year of simultaneous TOPEX/Poseidon and ERS-1 data. Negative supply rates reflect net local NO_3^- export in the particular year. Supply is positive (as is new production) when averaged over larger regions (Table 1) or longer periods. A band of high new production ($>0.5 \text{ mol N m}^{-2} \text{ yr}^{-1}$) extends from the Gulf Stream region, with its large EKE values, far into the much quieter eastern basin. In the subpolar North Atlantic, deep winter mixing brings large amounts of nitrate to the surface. High rates of nitrate supply are also associated with the upwelling regions off west Africa and at the Equator. We note, however, that despite the model's capacity to simulate eddies, nitrate fluxes into the southern part of the subtropical gyre are very low ($<0.01 \text{ mol N m}^{-2} \text{ yr}^{-1}$) and are, in fact, similar to supply rates reached by previous models that had a much coarser resolution¹⁹.

A meridional transect, zonally averaged from 50° to 30°W, reveals the eddies' contribution to new production (Fig. 2b). Between 32° and 45°N, nitrate fluxes of experiment A exceed those of the no-eddy run V by $\sim 0.3 \text{ mol N m}^{-2} \text{ yr}^{-1}$. In the subtropics and mid-latitudes, eddies account for 30–40% of the simulated new production (Table 1) which is consistent with previous idealized model studies⁶. We note, however, that all three experiments performed show an approximately zonal band, centred at $\sim 22^\circ \text{N}$, where new production is less than $0.01 \text{ mol N m}^{-2} \text{ yr}^{-1}$.

Close to Bermuda (32°N, 64°W), simulated new production increases from 0.37 (run V) to 0.50 and $0.53 \text{ mol N m}^{-2} \text{ yr}^{-1}$ in experiments C and A, respectively. The last two values agree well with geochemical estimates of new production². Interestingly, our model results are simultaneously consistent with much lower ($0.01 \text{ mol N m}^{-2} \text{ yr}^{-1}$) physical and biological estimates of new production further east³ (near 28°N, 23°W) in the subtropical gyre, indicating that even grossly different estimates of new production in the subtropical gyre need not be mutually inconsistent.

The simple ecosystem model used produces approximately similar ratios of new to total primary production in the various biological provinces (Table 1). Consequently, the pattern of primary production (Figs 1c, 2c) is well correlated with that of nitrate supply to the euphotic zone. Simulated primary production rates are $\sim 150 \text{ g C m}^{-2} \text{ yr}^{-1}$ in mid-latitudes and $100 \text{ g C m}^{-2} \text{ yr}^{-1}$ in the subpolar North Atlantic. Maximum values of up to $400 \text{ g C m}^{-2} \text{ yr}^{-1}$ are found in the upwelling region off west Africa, whereas $150 \text{ g C m}^{-2} \text{ yr}^{-1}$ are typical for the equatorial upwelling. Although these figures fall in between estimates compiled from *in situ* measurements²⁰ and those derived from ocean colour data^{21–23}, simulated primary production rates of less than $1 \text{ g C m}^{-2} \text{ yr}^{-1}$ in the southern part of the subtropical gyre are probably too low by at least an order of magnitude. In this region, our effort to account adequately for

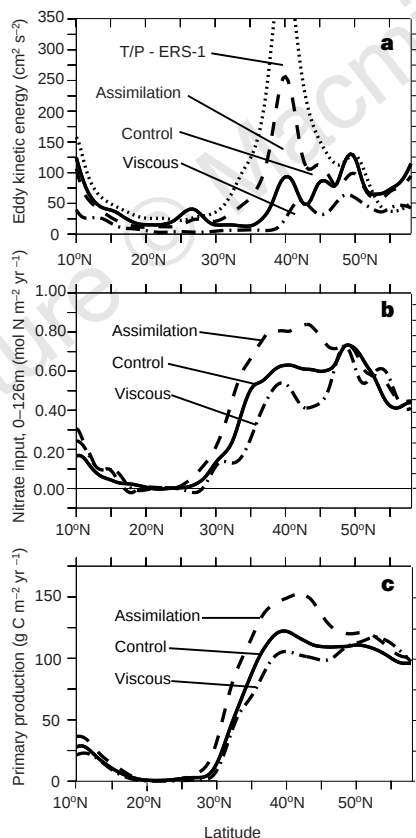


Figure 2 Meridional section, averaged from 50° to 30°W. **a**, Surface EKE (in $\text{cm}^2 \text{ s}^{-2}$) for the three numerical experiments as well as for the TOPEX/Poseidon-ERS-1 altimeter data. To allow for direct comparison, both observed and simulated EKE values have been obtained in the same way using geostrophic surface velocities derived from five-daily maps of sea surface height. **b**, The corresponding simulated nitrate flux into the upper 126 m (in $\text{mol N m}^{-2} \text{ yr}^{-1}$). **c**, Simulated annual primary production (in $\text{g C m}^{-2} \text{ yr}^{-1}$).

Table 1 Eddy kinetic energy, nitrate input and primary production

Region	Area (10 ¹² m ²)	Exp.	EKE* (cm ² s ⁻²)	Nitrate input (mol m ⁻² yr ⁻¹)	PP† (g C m ⁻² yr ⁻¹)	Nitrate input: PP
North Atlantic 8°–65° N	37	V	16	0.26	52	0.39
		C	74	0.29	57	0.40
		A	75	0.34	65	0.41
		Alt.	112			
Subpolar North Atlantic 50°–65° N	6.4	V	11	0.38	86	0.35
		C	28	0.39	87	0.36
		A	25	0.41	92	0.35
		Alt.	38			
Mid-lat. North Atlantic 30°–50° N	12.5	V	11	0.40	76	0.42
		C	87	0.48	91	0.42
		A	97	0.58	106	0.43
		Alt.	172			
Subtropical gyre 12°–30° N, 70°–22° W	8.2	V	4	0.016	3.1	0.41
		C	26	0.021	5.2	0.32
		A	21	0.047	7.5	0.49
		Alt.	38			
Canary basin 8°–30° N, 22° W-coast	1.8	V	8	0.47	98	0.38
		C	17	0.43	94	0.36
		A	18	0.43	95	0.36
		Alt.	37			
Gulf of Mexico 18°–32° N, 100°–82° W	1.9	V	21	0.22	44	0.39
		C	110	0.27	55	0.39
		A	166	0.37	69	0.42
		Alt.	270			
Equatorial Atlantic 8° S–8° N	10.5	V		0.40	60	0.53
		C		0.36	63	0.45
		A		0.37	66	0.45

Abbreviations used: EKE, eddy kinetic energy; PP, primary production; Alt., altimetric data. Experiments: V, viscous; C, control; A, assimilation.

* Model results and combined TOPEX/Poseidon and ERS-1 data are treated identically by computing EKE from five-daily sea-surface height maps.

† Diagnosed from the nitrogen-based model assuming a constant ratio of C:N=6.6.

mesoscale activity has simply resulted in an underprediction of primary production, similar to coarse-resolution models¹⁹.

There are at least three possible explanations for our model's failure to sustain reasonable primary production rates in the subtropical ocean. The ratio of new to total primary production may be much lower than is the case in the model configuration²⁴. Indeed, work in progress shows that by adding a fast recycling loop, subtropical primary production can be increased locally by a factor of 20 with only small changes of net nitrate supply and higher-latitude primary production. Primary production in the subtropical gyre might also be fuelled by lateral input of dissolved organic matter²⁵, which is not included in the present model. Our results suggest that Ekman drift or mesoscale eddies do not transfer sufficient amounts of nitrate far enough into the subtropical gyre, but more slowly overturning dissolved organic matter could accomplish much greater distances. Alternatively, one may ask to what extent the simulated nitrate supply to the euphotic zone is realistic. From the experiments compared above, we do not expect that even a possible doubling of EKE levels by future high-resolution models will generate order-of-magnitude changes in new production. However, recent estimates of pelagic nitrogen fixation in the subtropical North Atlantic²⁶ are as high as 0.07 mol N m⁻² yr⁻¹. If there were no other limiting nutrients, nitrogen fixation could locally account for an order-of-magnitude increase of new production.

The overall effect of increasing the level of eddy activity in this basin-wide high-resolution coupled ecosystem–circulation model is an increase of the annual new production from 0.74 Gt C yr⁻¹ (calculated between 8° and 65° N, and assuming C:N=6.6) in the viscous experiment V to 0.85 Gt C yr⁻¹ in the eddy-permitting run C to 1.0 Gt C yr⁻¹ in experiment A, in which the assimilation of altimeter data taken from the combined TOPEX/Poseidon and ERS-1 missions yields the most realistic representation of oceanic eddy activity that we can presently achieve. We find that eddies can not bridge the gap between grossly different estimates of new production in the subtropical gyre^{2,3}, but that the apparent dis-

crepancies may be consistently explained by different locations of the measurements taken. Our results also indicate that there remain large parts in the subtropical gyre where eddy activity is not sufficient to maintain observed levels of primary production. Alternative mechanisms that can provide the nitrogen supply need to be considered in future work. □

Received 17 December 1997; accepted 30 April 1998.

1. Eppley, R. W. & Peterson, B. J. Particulate organic matter flux and planktonic new production in the deep ocean. *Nature* **282**, 677–680 (1979).
2. Jenkins, W. J. Nitrate flux into the euphotic zone near Bermuda. *Nature* **331**, 521–523 (1988).
3. Lewis, M. R., Harrison, W. G., Oakey, N. S., Herbert, D. & Platt, T. Vertical nitrate fluxes in the oligotrophic ocean. *Science* **234**, 870–873 (1986).
4. Dugdale, R. C. & Goering, J. J. Uptake of new and regenerated forms of nitrogen in marine production. *Limnol. Oceanogr.* **12**, 196–206 (1967).
5. Falkowski, P. G., Ziemann, D., Kolber, Z. & Bienfang, P. K. Role of eddy pumping in enhancing primary production in the ocean. *Nature* **352**, 55–58 (1991).
6. McGillicuddy, D. J. Jr & Robinson, A. R. Eddy-induced nutrient supply and new production in the Sargasso Sea. *Deep-Sea Res.* **144**, 1427–1450 (1997).
7. Dadou, I., Garçon, V., Andersen, V., Flierl, G. R. & Davis, C. S. Impact of the North Equatorial Current meandering on a pelagic ecosystem: A modeling approach. *J. Mar. Res.* **54**, 311–342 (1996).
8. Williams, R. G. & Follows, M. J. The Ekman transfer of nutrients and maintenance of new production over the North Atlantic. *Deep-Sea Res.* **145**, 461–489 (1998).
9. Oschlies, A. & Garçon, V. An eddy-resolving coupled physical-biological model of the North Atlantic. Part I: Sensitivity to advection numerics and mixed layer physics. *Glob. Biogeochem. Cycles* (submitted).
10. Bryan, F. O. & Holland, W. R. in *Parameterization of Small-Scale Processes* (eds Müller, P. & Henderson, D.) 99–115 (Proc. 'Aha Huliko'a, Hawaiian Winter Workshop, Spec. Publ., Hawaii Inst. Geophys, Manoa, 1989).
11. Döschner, R., Böning, C. W. & Herrmann, P. Response of circulation and heat transport in the North Atlantic to changes in thermohaline forcing in northern latitudes: a model study. *J. Phys. Oceanogr.* **24**, 2306–2320 (1994).
12. Blanke, B. & Delecluse, P. Variability of the tropical Atlantic Ocean simulated by a general circulation model with two different mixed-layer physics. *J. Phys. Oceanogr.* **23**, 1363–1388 (1993).
13. Paulson, C. A. & Simpson, J. J. Irradiance measurements in the upper ocean. *J. Phys. Oceanogr.* **7**, 952–956 (1977).
14. Levitus, S., Conkright, M. E., Reid, J. L., Najjar, R. G. & Mantyla, A. Distribution of nitrate, phosphate and silicate in the world oceans. *Prog. Oceanogr.* **31**, 245–273 (1993).
15. Lafore, J.-P. et al. The Meso-NH atmospheric simulation system. Part I: Adiabatic formulation and control simulations. *Ann. Geophysicae* **16**, 90–109 (1998).
16. Stammer, D. & Böning, C. W. Mesoscale variability in the Atlantic Ocean from GEOSAT Altimetry and WOCE high resolution numerical modeling. *J. Phys. Oceanogr.* **22**, 732–752 (1992).
17. Le Traon, P. Y., Gaspar, P., Bouysse, F. & Makhmara, H. Using TOPEX/POSEIDON data to enhance ERS-1 data. *J. Atmos. Ocean. Technol.* **12**, 161–170 (1995).
18. Oschlies, A. & Willebrand, J. Assimilation of Geosat altimeter data into an eddy-resolving primitive equation model of the North Atlantic Ocean. *J. Geophys. Res.* **101**(C6), 14175–14190 (1996).

19. Sarmiento, J. L. *et al.* A seasonal three-dimensional ecosystem model of nitrogen cycling in the North Atlantic euphotic zone. *Glob. Biogeochem. Cycles* **7**, 417–450 (1993).
20. Berger, W. H. in *Productivity of the Ocean: Present and Past* (eds Berger, W. H., Smetacek, V. S. & Wefer, G.) 429–455 (Wiley, Chichester, 1989).
21. Antoine, D., André, J.-M. & Morel, A. Oceanic primary production 2. Estimation at global scale from satellite (coastal zone color scanner) chlorophyll. *Glob. Biogeochem. Cycles* **10**, 57–69 (1996).
22. Sathyendranath, S., Longhurst, A., Caverhill, C. M. & Platt, T. Regionally and seasonally differentiated primary production in the North Atlantic. *Deep-Sea Res.* **42**, 1773–1802 (1995).
23. Behrenfeld, M. J. & Falkowski, P. G. Photosynthetic rates derived from satellite-based chlorophyll concentration. *Limnol. Oceanogr.* **42**, 1–20 (1997).
24. Platt, T. & Harrison, W. G. Biogenic fluxes of carbon and oxygen in the ocean. *Nature* **318**, 55–58 (1985).
25. Archer, D., Pelzer, E. T. & Kirchman, D. L. A timescale for dissolved organic carbon production in equatorial Pacific surface waters. *Glob. Biogeochem. Cycles* **11**, 435–452 (1997).
26. Gruber, N. & Sarmiento, J. L. Global patterns of marine nitrogen fixation and denitrification. *Glob. Biogeochem. Cycles* **11**, 235–266 (1997).
27. Sambrotto, R. N. *et al.* Elevated consumption of carbon relative to nitrogen in the surface ocean. *Nature* **363**, 248–250 (1993).

Acknowledgements. We thank J. F. Minster for discussions, and W. Koeve and G. Evans for comments. This work was supported through the European Union ESCOBA program and the German JGOFS project funded by the BMBF.

Correspondence and requests for materials should be addressed to A.O. (e-mail: aoschlies@ifm.uni-kiel.de).

Local and global vectors in desert ant navigation

M. Collett^{*}, T. S. Collett[†], S. Bisch[‡] & R. Wehner[§]

^{*}BBSRC-NERC Ecology and Behaviour Group, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK

[†]Sussex Centre for Neuroscience, School of Biological Sciences, University of Sussex, Brighton BN1 9QG, UK

[‡]Zoologisches Institut, Universität Bonn, Kirschallee 1, 53115 Bonn, Germany

[§]Department of Zoology, University of Zürich, Winterthurerstrasse 190, CH 8057 Zürich, Switzerland

Desert ants returning from a foraging trip to their nest navigate both by path integration and by visual landmarks^{1–3}. In path integration, ants compute their net distance and direction from the nest throughout their outward¹ and return⁴ journeys, and so can always return directly home from their current location¹. As the path-integration vector is updated over the entire journey, we call it a global vector. On a familiar route, when ants can steer by visual landmarks, they adopt a fixed and often circuitous path consisting of several separate segments that point in different directions^{2,3,5}. Here we show that, as in honeybees^{6–8}, such multi-segment journeys are composed partly of stored local movement vectors, which are associated with landmarks and are recalled at the appropriate place. We also show that a local vector learnt at one value of the global vector can be recalled at many values, and that expression of the global vector is temporarily inhibited while the local vector is used. These results indicate that the global vector is ignored during navigation through familiar, cluttered territory, but that it re-emerges to take the ant home once the insect leaves the clutter and other guidance strategies cease to operate.

Insects steer by familiar landmarks both when homing normally from a feeding site and if transported there from their nest⁹, using a variety of guidance strategies³. Indeed, the complex homeward paths of desert ants navigating through scrub are almost identical under these two conditions³, even though the predicted values of the global vector differ greatly. This similarity in homeward routes raises the question of what happens to the global vector during landmark navigation. To find out, we have examined interactions between global and local vectors by presenting ants with experimental situations in which the two predicted vectors point in different directions (Fig. 1).

Desert ants (*Cataglyphis fortis*) were trained on a flat, sandy area along a two-leg route to and from a feeder (Fig. 1A). The first leg from the nest was over 8 m of open ground. The second leg was

perpendicular to the first along a narrow channel 8 m in length that allowed a view of the sky but not of the surrounding landscape¹⁰. To avoid complications that might be introduced if the channel itself were to be used as a landmark by approaching or departing ants, we hid the channel in a trench, making it invisible to ants more than a few centimetres away. Ants collected food in a compartment at the end of the channel and then returned east along the channel to its exit, from which they walked south to reach their nest.

Trained ants were caught either at the feeding box, where their global vector was directed 11.3 m southeast towards the next (Fig. 1B), or at the end of the return trip to the nest, where the global vector was zero (Fig. 1C). Each ant was then carried to a test area, and released into a feeding box at the end of another channel which was 2, 4, 8 or 11 m long. Here it found a biscuit crumb which it carried along the channel and then homewards across open ground. The value of the global vector on exiting the test channel depends on where the ant was caught and on the length and orientation of the channel (Fig. 1 and Table 1). The postulated local vector, on the other hand, always points south (Fig. 1).

Ants taken at the nest and placed in an east-pointing channel tended to walk in their accustomed southerly direction on reaching open ground (Fig. 2d–f and Table 1). This southward path was not driven by the global vector, for the vector was zero at the release site and increased in a westerly direction as the ant walked east along the channel (Fig. 1C). Once out of the channel, there were no local visual landmarks to guide the ants. We suggest that their path was

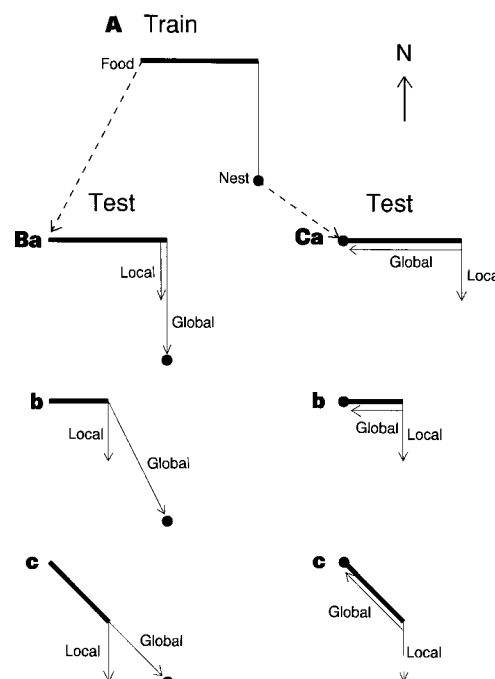


Figure 1 Rationale of experiments. **A**, Training arrangement. Dot indicates nest. Thick line shows the channel which forced the ants to travel west to reach food at the end of the channel and to return east on the first leg of their homeward journey. Thin line shows directed path over open ground. **B**, **C**, Dot illustrates position of the fictive nest, where the predicted value of the global vector is zero. Thin arrows show the directions of the predicted global and local vectors at the exit to the channel. **B** (left), ants are taken from the training feeder (left dotted arrow from **A**). **a**, Test condition reproducing training, when directions of local and global vectors coincide. **b**, **c**, A short and/or rotated test channel was used; global and local vectors point in different directions. **C**, Ants are captured at the nest and taken to the test channel (right dotted arrow from **A**). Predicted directions of vectors in tests performed when ants are placed at the end of the channel after capture at the nest. The fictive nest is then at the release site in the channel so that global vectors point towards the release site, whereas local vectors point south.

dictated by a local south-pointing vector, which ants had associated with the channel and retrieved on leaving it. The prominence of the local vector depended on the length of the channel. Trajectories on exiting the 2-m channel (Fig. 2f) were shorter than those on leaving the 8-m channel (Fig. 2d) ($P < 0.05$, Mann–Whitney test) and were less well directed.

Is the local vector controlled by compass cues, or did ants simply make a 90° turn to the right on leaving the channel? We tested ants in southeast- and northeast-pointing channels (Fig. 3). Ants that had short global vectors when they left the channel tended to walk southward (Table 1). Thus, the trajectory directions of ants caught at the nest and released into the 2-m southeast-pointing channel (Fig. 3f) or the 4-m northeast-pointing channel (Fig. 3g), or taken from the feeder to the 8-m southeast-pointing channel (Fig. 3a), were significantly different from the global vector and the 90° turn ($P < 0.01$), but were not different from the local vector ($P > 0.05$, 2-m southeast and 4-m northeast channels; $P > 0.01$, 8-m southeast channel). The direction of the local vector was thus determined by compass (probably skylight³) information. However, local vectors can be less salient if the global vector is longer and the direction of the test channel differs from that in which ants were trained (compare Fig. 2d and Fig. 3d). In a control test, no directional preferences were seen in the trajectories of ants caught at the nest

and released on open ground (Fig. 3h). It seems that ants have associated a compass-bound, local vector with a familiar channel, and that they retrieved the local vector when walking along and exiting that channel.

Whenever local vectors were performed, they were substantially shorter than the distance covered by ants homing in the training configuration (Table 1). Was the local vector short because it carried the ant to small, local features on the training ground that the ant uses as landmarks? To minimize the use of such features, we placed a large (44.5 cm diameter, 80.5 cm high) black cylinder just south of the nest. The ants used the cylinder immediately after they left the channel, zigzagging towards it. After 5 days foraging with this cylinder in place, ants were taken at their return to the nest and released in the end of the 4-m east-pointing channel on the test ground where there was no cylinder. Trajectories were neither detectably shorter nor longer than before (Table 1). The lack of any shortening indicates that the vector probably does not encode the distance between the ant's leaving the channel and its coming under the influence of the cylinder. On the other hand, the absence of an increase in length suggests that the local vector does not encode the distance between leaving the channel and reaching the cylinder (assuming that the subsequent lack of a cylinder on the test ground does not shorten the vector). Instead, the role of the local

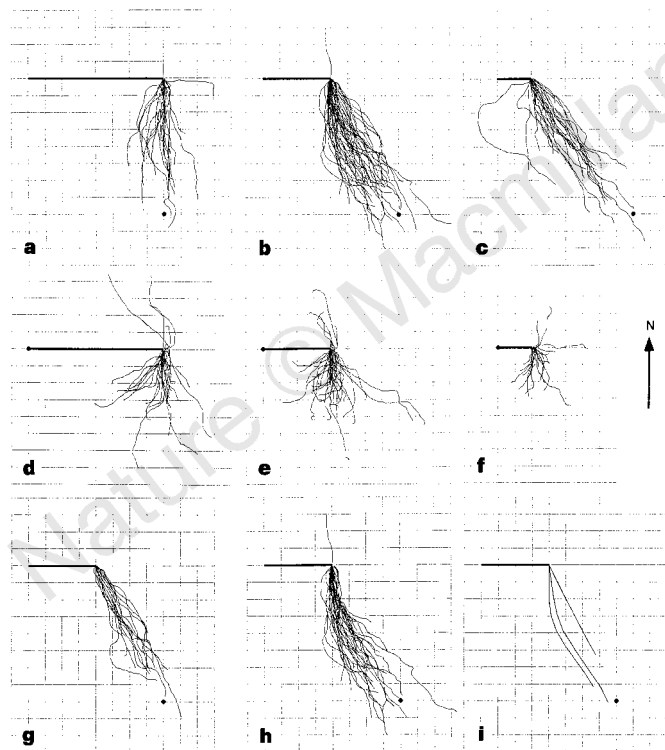


Figure 2 Trajectories of ants 'homing' on the test area after their release at the end of a 2-, 4- or 8-m east-pointing channel. The ants' paths were recorded from leaving the channel until they had changed direction several times, indicating searching behaviour¹². Analysis and figures include trajectories until the initiation of search behaviour. **a–c**, Data from ants collected at feeder. **d–f**, Data from ants collected at nest. **g, h**, Trajectories in **b** are segregated according to whether ants went to the left (**g**) or to the right (**h**) of a point 2 m south and 0.5 m east of the channel exit. **i**, Means of the trajectories of **b, g** and **h** computed by a 'wave-front' method. The mean direction of the first 2 m of the trajectories shown in **h** relative to south (0°) (mean = 3° E, s.d. = 6.8° , $n = 37$) is significantly different from that of the same trajectories between 4 m and 6 m from the channel exit (mean = 30° E, s.d. = 13.0° , $n = 27$) ($P < 0.001$, Watson–Williams test). The origin of the global vector (position of fictive nest) is marked with a black circle. Grid lines are 1 m apart.

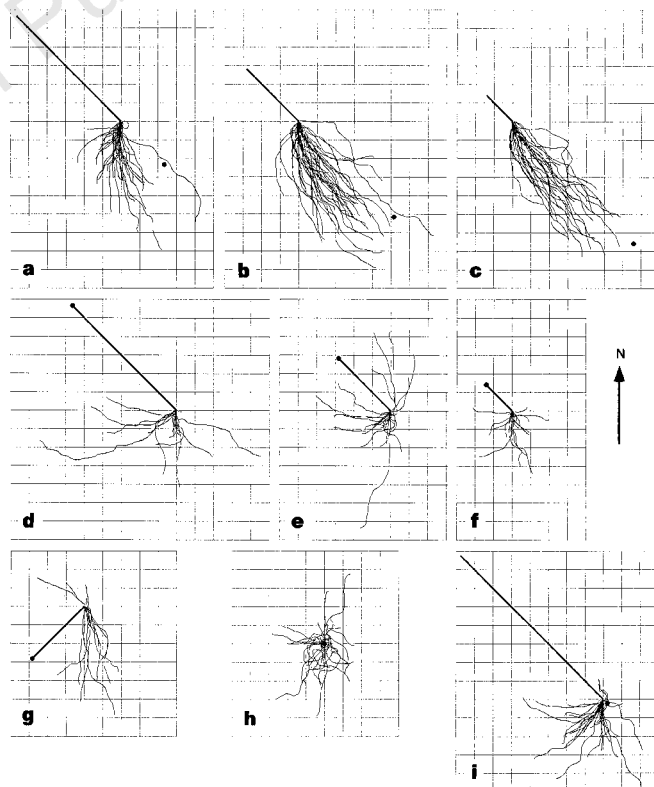


Figure 3 Trajectories of ants on the test area after their release at the end of a southeast- or northeast-pointing channel or on open ground. **a–f**, As for Fig. 2 but with channel orientated southeast. **g**, Trajectories from ants collected at nest and released into a 4-m northeast-pointing channel. **h**, Trajectories of ants collected at nest and released on melon on open ground. **i**, Trajectories of ants collected at feeder after their release into an 11-m southeast-pointing channel.

Table 1 Directions and distances of trajectories

Corresponding figure	Capture site	Channel	Direction with respect to south (°)		Distance (m)		Number of observations
			Predicted global	Observed	Predicted global	Observed	
2a	Feeder	E 8 m	0	1W ± 11	8.0	5.77 ± 0.33	20
2b	Feeder	E 4 m	27E	21E ± 8	8.94	7.48 ± 0.30	57
2c	Feeder	E 2 m	37E	31E ± 7	10.0	8.58 ± 0.33	30
2d	Nest	E 8 m	90W	2W ± 45**	8.0	2.18 ± 0.21	34
2e	Nest	E 4 m	90W	6E ± 36**	4.0	1.91 ± 0.15	57
2f	Nest	E 2 m	90W	27E ± 43	2.0	1.55 ± 0.14	30
3a	Feeder	SE 8 m	45E	8W ± 21*	3.31	2.06 ± 0.21	35
3b	Feeder	SE 4 m	45E	26E ± 14	7.31	5.87 ± 0.31	42
3c	Feeder	SE 2 m	45E	35E ± 10	9.31	7.32 ± 0.32	30
3d	Nest	SE 8 m	135W	35W ± 40	8.0	1.67 ± 0.37	17
3e	Nest	SE 4 m	135W	23W ± 56**	4.0	1.39 ± 0.21	28
3f	Nest	SE 2 m	135W	4E ± 32***	2.0	1.18 ± 0.14	25
3g	Nest	NE 4 m	45W	8E ± 7***	4.0	2.49 ± 0.34	15
3i	Feeder	SE 11 m	45E	20W ± 39	0.31	2.21 ± 0.27	38
Text	Nest	E 4 m with cylinder	90W	19W ± 46*	4.0	2.39 ± 0.24	36

'Predicted global' columns show calculated values of the global vector at the exit to the test channel. 'Observed' columns show means and standard deviations of the vector between the channel exit and the trajectory endpoint (where the ant begins search behaviour¹¹). Single, double and triple asterisks indicate cases in which ants used local vectors. The means of the observed directions do not differ from that of the local vector (* $P > 0.01$; **, *** $P > 0.05$), but do differ from the direction of the global vector ($P < 0.01$) and also, when applicable, from the perpendicular to the channel (*** $P < 0.01$).

vector may be to ensure that the ant sets off in the correct direction along a path segment, allowing the ant to view the next set of landmarks from a standard vantage point and orientation¹¹.

Are local vectors found only in ants that have been confined to channels or are they also associated with other location-specific cues? We trained ants from a second nest to feed at a site 13 m south of their nest along a straight route that was flanked on either side by a row of three black cylinders (Fig. 4a). Trained ants were caught

either at the feeder or after reaching their nest, and taken to a test area where there was another corridor of cylinders. Ants collected at the feeder and released at the corridor entrance walked down the corridor and continued in the same direction for a few metres after passing the last cylinder. Ants with zero global vector followed a search path¹² before discovering the corridor. They then walked through the corridor and their trajectories also continued beyond its end (Fig. 4b).

The crucial test to detect a compass-based local vector was to shift the orientation of the corridor through 45° (Fig. 4c). Ants caught at the nest were released with zero global vector in a northwest-pointing corridor. After an initial search, most ants (10 out of 17) travelled northwest along the corridor and, on leaving it, turned consistently north. The remainder turned north earlier as though they had retrieved their local vector after passing only one or two of the landmarks. Thus, despite the reorientation of the landmarks, the paths of all the ants eventually pointed north (Fig. 4d). In contrast, ants collected at the nest and released on open ground, clear of cylinders, searched for a long period with no directional preference (data not shown). The ants must have associated a northerly directed local vector with the landmarks.

How do global and local vectors interact? For ants accustomed to the two-leg route, global and local vectors on exiting the test channel always pointed in different directions from each other, apart from in the test condition reproducing training (Fig. 2a). In many cases ants resolved this conflict by transiently inhibiting the expression of the global vector (Table 1). This suppression is best seen with ants taken at the feeder and released in the 4-m east-pointing channel (Fig. 2b). Some trajectories (33%) were directed entirely southeast along the global vector (Fig. 2g), but most trajectories (66%) first pointed south along the local vector and then turned into the direction of the global vector (Fig. 2h, i); the global vector emerges only after the local vector has been performed. Some ants caught at the feeder and released in 2-m east-pointing, 4-m southeast-pointing, or 2-m southeast-pointing channels behaved similarly (Figs 2c and 3b, c), but the proportion doing so was smaller (30%, 47% and 30%, respectively). Local and global vectors are thus not simply combined; instead, expression of the global vector can be inhibited while the local vector is performed. The global vector is probably continually updated while it is suppressed, so that its value is always appropriate for guiding the ant home⁴.

A variant of the interaction between local and global vectors is sometimes found in ants released with zero global vector. For ants taken after their return to the nest, the release site at the end of the channel became the origin of the global vector, and ants often tended to return there. This tendency was very marked in ants

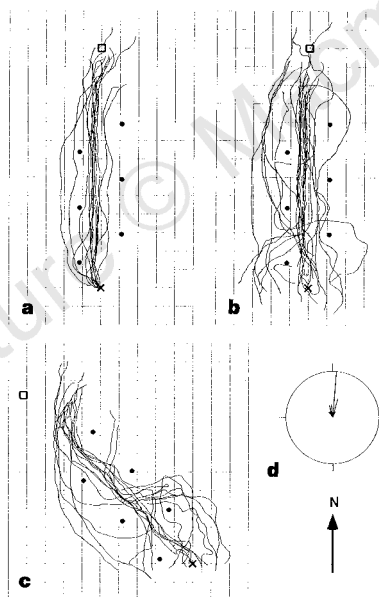


Figure 4 Trajectories of ants trained to a feeder at the end of a corridor of black cylinders (20 cm in diameter and of different heights ranging from 20 cm to 53 cm). **a, b**, Tests with landmarks arranged in training situation. Ants were caught either at the feeding site (**a**) or at the nest (**b**). Mean distance of vector from release site to beginning of search is 12.47 m (s.d. = 0.92 m, $n = 14$) for ants caught at the feeder, and 11.82 m (s.d. = 0.59 m, $n = 22$) for ants caught at the nest. The initial search path of ants with zero global vector is not shown; a cross marks the release site to the south; a box marks the position of fictive nest relative to the corridor of cylinders. **c**, Tests with landmarks rotated through 45° for ants caught at the nest. The direction of the ants just before leaving the corridor of landmarks (means = 44°W, s.d. = 4.3°, $n = 9$). **d**, Circular histogram of the directions of the last 3 m of all 17 trajectories shown in **c**. Using the 3-m circle centred at the endpoint of the trajectory, direction is defined as the inward radial that starts at the intersection of the trajectory with the circle.

exiting the 8-m or 4-m southeast-pointing test channels (Fig. 3d, e). The direction of turning back was not random but was biased to the right ($P < 0.001$, binomial test), indicating that, even here, the local vector influences the initial direction that the ant takes when it follows the global vector.

Inhibiting the expression of global vectors allows local vectors to be executed without interference. The advantage this brings can be seen when desert ants take complex, fixed routes through familiar scrub, where there can be sizeable discrepancies between the optimal path between bushes and the direction indicated by the global vector³. Our results indicate that the ant's path is then governed predominantly by a mixture of landmark-based guidance mechanisms and stored local vectors, with little contribution from the global vector. This suppression of the global vector by local vectors helps to explain the provocative results with which we began: that an ant's normal homeward path, as it weaves through familiar scrub, is identical to the path it takes when it is displaced to the same feeding site with zero global vector³. □

Methods

Training and testing procedure. Ants were free to collect watermelon and sometimes biscuit crumbs in a feeding compartment at the end of the 6-cm-wide training channel during testing and for 2 days before. Ants were tested singly at an open test area ~100 m west of the training ground. A 1-m grid was painted on the ground so the ant's path could be recorded by an observer drawing on squared paper. Three separate trenches were dug into the test field and channels¹⁰ placed in them: one pointed east, as did the training channel, and the two others pointed southeast and northeast. The lengths of the test channels within the trenches could be varied. A trained ant was caught either at the feeder in the channel or at the end of its return trip within 50 cm of the nest entrance. It was carried in a darkened container to the test area where it was placed onto a piece of melon in a feeding box at the end of a channel, and provided with a biscuit crumb which it picked up and carried homewards.

Computing mean trajectories. Mean trajectories (Fig. 2i) were computed in 1-m steps. For the first step we computed for each trajectory the mean direction of the vector connecting the channel exit to the intersection of each trajectory with a circle of 1 m radius centred on the exit. The position of the mean direction on the circle became the origin of step 2. The mean vector from the channel exit to the circle was projected 1 m beyond the origin of step 2. We then computed for each trajectory the vector from the origin of step 2 to where the trajectory crossed the normal to the tip of the projected mean vector. The intersection of the mean direction of these vectors with the normal defined the origin of step 3. This process was repeated until fewer than ten trajectories contributed to the mean.

Received 5 November 1997; accepted 16 April 1998.

1. Wehner, R. & Wehner, S. Insect navigation: use of maps or Ariadne's thread. *Ethol. Ecol. Evol.* **2**, 27–48 (1990).
2. Collett, T. S., Dillmann, E., Giger, A. & Wehner, R. Visual landmarks and route following in desert ants. *J. Comp. Physiol. A* **170**, 435–442 (1992).
3. Wehner, R., Michel, B. & Antonson, P. Visual navigation in insects: coupling of egocentric and geocentric information. *J. Exp. Biol.* **199**, 129–140 (1996).
4. Schmidt, I., Collett, T. S., Dillier, F.-X. & Wehner, R. How desert ants cope with enforced detours on their way home. *J. Comp. Physiol. A* **171**, 285–288 (1992).
5. Rosengren, R. Route fidelity, visual memory and recruitment behaviour in foraging wood ants of the genus *Formica* (Hymenoptera, Formicidae). *Acta Zool. Fenn.* **133**, 1–106 (1971).
6. Collett, T. S. & Baron, J. Learnt sensori-motor mappings in honeybees: interpolation and its possible relevance to navigation. *J. Comp. Physiol. A* **176**, 287–298 (1995).
7. Menzel, R., Geiger, K., Joerges, J., Müller, U. & Chittka, L. Bees travel novel homeward routes by integrating separately acquired vector memories. *Anim. Behav.* **55**, 139–152 (1998).
8. Srinivasan, M. V., Zhang, S. W. & Bidwell, N. J. Visually mediated odometry in honeybees. *J. Exp. Biol.* **200**, 2513–2522 (1997).
9. Baerends, G. P. Fortpflanzungsverhalten und Orientierung der Grabwespe *Ammophila campestris*. *Tijdschr. Entomol.* **84**, 68–275 (1941).
10. Müller, M. & Wehner, R. Path integration in desert ants, *Cataglyphis fortis*. *Proc. Natl Acad. Sci. USA* **85**, 5287–5290 (1988).
11. Collett, T. S. & Rees, J. A. View-based navigation in hymenoptera: multiple strategies of landmark guidance in the approach to a feeder. *J. Comp. Physiol. A* **181**, 47–58 (1997).
12. Wehner, R. & Srinivasan, M. V. Searching behaviour of desert ants, genus *Cataglyphis* (Formicidae, Hymenoptera). *J. Comp. Physiol.* **142**, 315–338 (1981).

Acknowledgements. We thank F.-X. Dillier for help and P. Graham for comments. Financial support came from the BBSRC, the Human Frontiers Science Program and the Swiss National Science Foundation.

Correspondence and requests for materials should be addressed to T.S.C. (e-mail: t.s.collett@sussex.ac.uk).

Species extinction and the relationship between distribution and abundance

C. N. Johnson

Department of Zoology and Tropical Ecology, James Cook University, Townsville, Queensland 4811, Australia

Within taxonomic groups, there is almost always a positive relationship between the size of geographic range and the local abundance of species^{1–4}. This pattern has attracted much interest, and several ecological mechanisms have been proposed as causes of it⁵. However, these hypotheses do not consider the effect of the extinction of rare species on range-abundance relationships. If both range size and local abundance influence the risk of extinction, species with small ranges might avoid extinction if they have high local abundance, whereas species with low local abundance might avoid extinction if they are widespread; species with both small range and low local abundance should be at high risk. This interaction between range, abundance and extinction should produce negative correlations between range and abundance in groups that have experienced many extinctions. Here I test this idea using Australian marsupials, and I show that although the relationship between range size and local abundance is positive for recently evolved species, it is negative for ancient species. This indicates that positive relationships between range size and abundance may be generated during adaptive radiation, but are then gradually reversed as a result of differential extinction.

I defined recently evolved species as those known to have radiated from a common ancestor within the last 4 million years, and ancient species as those that diverged from their closest extant relative more than 4 million years ago. Analysis of the relationship between range

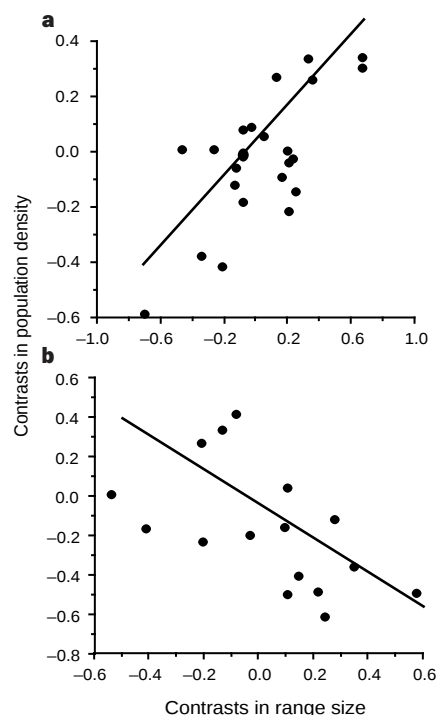


Figure 1 Relationships between phylogenetically independent contrasts²⁷ in geographic range size and population density for **a**, recently evolved marsupials ($r = 0.72$, $n = 23$, $P < 0.001$), and **b**, ancient species ($r = -0.56$, $n = 16$, $P < 0.05$).

and abundance among extant species from both groups showed a strong positive relationship among recently evolved species, as expected^{1–5}, but a strong negative relationship among ancient species (Fig. 1).

In theory, local abundance and size of geographic range should contribute independently to extinction risk. Local abundance is thought to be negatively related to the probability of local extinction resulting from demographic and environmental stochasticity^{6–8}, whereas range size is negatively related to the probability of extinction from loss of habitat⁹. These two causes of extinction will be independent to the extent that local stochasticity is unrelated to long-term trends in the distribution of habitat. Such independence should have two effects on patterns of extinction: first, extinction risk will be compounded in species that have both small ranges and low abundance; and second, range size and abundance may compensate for one another in determining how prone a species is to extinction. In Fig. 2 I provide evidence of both effects in marsupials. The distribution of species through range-abundance space differs between recent and ancient marsupials because of the absence of ancient marsupial species with small ranges and low densities. We can assume that such species once existed as relatives of the surviving ancient marsupials, because there are so many species with small range and low density in the recent fauna, and that they have become extinct as a result of their rarity. Compensation is implied by the fact that the lower limits of range size and local abundance are negatively related among the surviving ancient marsupial species (Fig. 2b). This compensation seems to be exact, because the slope of the relationship between the lower limits of range size and local abundance is roughly -1 .

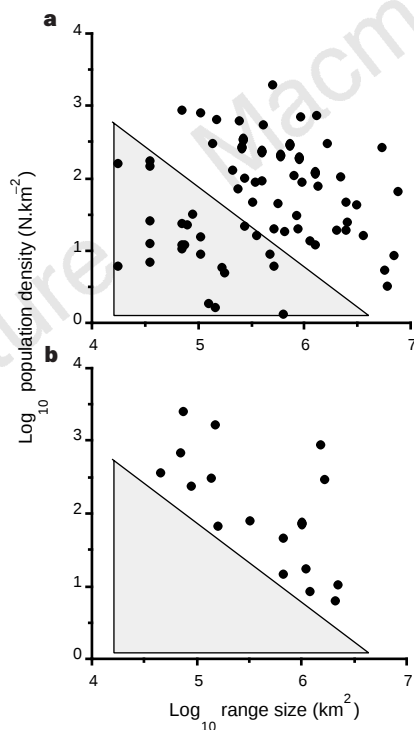


Figure 2 Relationships across species between geographic range size and population density of marsupial species for **a**, all species other than those placed in the 'ancient' category ($r = 0.01$, $n = 67$, n.s.), and **b**, species in the ancient category ($r = -0.65$, $n = 17$, $P < 0.005$). The shaded triangle identifies the region from which species would need to have gone extinct to have produced the relationship observed among ancient species. Its shape was determined by the slope of the lower limit of the cloud of points in **b**, which was estimated using the method described in ref. 29; the data set was divided into vertical strips of 0.25 units width each, and the lowest estimate of density in each strip was regressed on range size ($F_{1,8} = 121.89$, $P < 0.001$, $R^2 = 0.96$, slope = -1.12).

Several mechanisms, operating over long periods of time and at large spatial scales, could be involved in producing this compensation. For example, species with large ranges might be able to tolerate lower population densities (and hence more frequent extinction of local populations) because of a greater potential for migration from distant parts of the range to counter local extinctions. Conversely, species with high densities might persist despite small range size because high local abundance results in the production of large numbers of migrants which, especially at the range boundary, allows the tracking of shifting habitats. An alternative argument is that range size and local abundance do not act independently to determine extinction risk; rather, they might simply determine total population size, which in turn is directly related to the probability of extinction. In that case, the inverse relationship between the lower limits of range and abundance shown in Fig. 2b might reflect the minimum total population size necessary for long-term persistence of species. This interpretation seems implausible, because it implies that the effect of total population size on probability of extinction is the same regardless of whether the species is spread through a large or a small range; there is no single process of extinction that could be expected to act in this way. Species that have both large geographic range and high population densities should be the most resistant to extinction; this is consistent with the observation that such species are represented among both ancient and recent marsupials. This analysis is one of the first to attempt to disentangle the effects of range size and local abundance on extinction risk; the few previous studies that have examined this question also indicate that the effects of these variables may be independent¹⁰.

Characteristics of species other than range size and local abundance have been suggested to influence the likelihood of extinction. For example, large-bodied species might have lower risks of extinction than small-bodied species because their populations fluctuate less⁷, and species with high fecundity may have lower risks of extinction because they recover more quickly from population declines¹¹. Neither of these predictions was supported by comparison of ancient and other marsupials: the two groups did not differ in either mean annual fecundity ($t_{10} = 1.57$, not significant) or body size ($t_{13} = -0.09$, not significant; both comparisons based on phylogenetically independent contrasts). Instead, the patterns shown in Fig. 2 indicate the primacy of range size and local abundance, independent of other differences among species, in determining probability of extinction.

Both the positive relationship between range and abundance in recent marsupials and the opposite relationship among ancient marsupials were strong, but across the marsupials as a whole they blended to produce a weak positive relationship (using independent contrasts, $r = 0.24$, $n = 68$, $P = 0.05$). Distribution-abundance relationships are typically weak, especially when (as here) comparisons are made within higher taxonomic groupings and at broad spatial scales⁴. My results indicate that the vagueness of relationships observed in other groups may be due partly to the confounding effects of history on the form of the relationship, rather than to a lack of power in the underlying mechanisms. Accounting for the effects of taxon age may reveal stronger underlying patterns, and provide a mechanistic explanation for cases in which range and abundance are negatively related¹². □

Methods

I measured range sizes of marsupials extant on mainland Australia from published distribution maps¹³ by counting the number of equal-area (17,500 km²) grid cells overlapped by the range of each species, then multiplying these by the cell area to convert them to units of km². When species distributions contained disjunctions of more than 5° of latitude or longitude, the parts of the range were treated as separate units for analysis (in most cases where this was true, the distinct populations have subspecific status). I collected estimates of population density from the published literature and from

unpublished studies; data on body weights and fecundity were taken from ref. 13. All variables were log-transformed before analysis and where several estimates of density were available I used the mean of the log-transformed values. Data on population density were available for 81 species of the total of 161.

The phylogeny used in this analysis was based on a major synthesis¹⁴ of DNA-hybridization studies of Australian marsupials. This phylogeny resolves the relationships of most genera and some subgeneric groupings, and provides estimates of age for each node. Relationships among species within genera, and the positions of genera not included in ref. 14, were taken from refs 15–24. I used the CAIC program²⁵ to calculate phylogenetically independent contrasts in the variables of interest. This program implements Pagel's version of Felsenstein's method^{26,27} to remove the non-independence due to phylogenetic association from relationships between variables measured on species. This analysis requires information on branch lengths, which was not available for many of the branches in the phylogeny I used. In the absence of such information, CAIC offers two default assumptions: all branch lengths may be set equal (this is equivalent to assuming a punctuational mode of evolution) or branch lengths may be assigned on the assumption that the ages of taxa are in direct proportion to the number of species they contain. All analyses were repeated using both assumptions, but this made no significant difference to the results. For the analyses presented here I assumed equal branch lengths.

The cut-off between recent and ancient marsupials was set at 4 million years because this is approximately twice the mean species lifespan for mammals²⁸. This meant that beyond 4 million years ago, differential extinction could be expected to have a major structuring effect on mammalian assemblages. I was forced to use a cut-off rather than treat age as a continuous variable because the divergence dates for many species were not known precisely, but these species could be placed into one or other of the age categories on the basis of their position relative to nodes of known age. Those species that could be placed into one or other of the categories were as follows.

Recent species: all species of *Isodon*, *Macropus*, *Petrogale*, *Trichosurus* and *Lasiorehinus*, *Petaurus gracilis*, *P. norfolkensis*, *Thylacynus*, *T. stigmatica*, and *Pseudochirus cinereus*. Within this group, analysis by independent contrasts was completed only to nodes of age 4 million years or less.

Ancient species: *Acrobates pygmaeus*, *Aepyprymnus rufescens*, *Burramys parvus*, *Gymnobelideus leadbeateri*, *Hemibelideus lemuroides*, *Hypsiprymnodon moschatus*, *Lagostrophus fasciatus*, *Myrmecobius fasciatus*, *Petauroides volans*, *Petropseudes dahli*, *Phascogale cinereus*, *Pseudocheirus peregrinus*, *P. occidentalis*, *Setonix brachyurus*, *Tarsipes rostratus*, *Vombatus ursinus* and *Wyulda squamicaudata*. A date for the divergence of *Pseudocheirus peregrinus* and *P. occidentalis* was not available, but was assumed to be ancient because of the early separation of forested habitat in southeastern and southwestern Australia, and the fact that *Pseudocheirus* is thought to have diverged from the *Pseudochirus* group of ringtail possums more than 20 million years ago¹⁴.

To test the sensitivity of the analysis to the position of the cut-off, I re-analysed the data using a cut-off of 10 million years to distinguish ancient and recent species, but the correlations produced by this re-analysis ($r = 0.55$, $n = 35$, $P < 0.001$ for recent taxa and $r = -0.82$, $n = 10$, $P < 0.01$ for ancient taxa) did not substantively alter the result shown in Fig. 1.

Received 17 March; accepted 5 May 1998.

- Hanski, I. Dynamics of regional distribution: the core and satellite species hypothesis. *Oikos* **38**, 210–221 (1982).
- Brown, J. H. On the relationship between abundance and distribution of species. *Am. Nat.* **124**, 255–279 (1984).
- Lawton, J. H., Nee, S., Letcher, A. J. & Harvey, P. H. in *Large-scale Ecology and Conservation Biology* (eds Edwards, P. J., May, R. M. & Webb, N. R.) 41–58 (Blackwell Scientific, London, 1994).
- Gaston, K. The multiple forms of the interspecific abundance-distribution relationship. *Oikos* **76**, 211–220 (1996).
- Gaston, K. J., Blackburn, T. M. & Lawton, J. H. Interspecific abundance-range size relationships: an appraisal of mechanisms. *J. Anim. Ecol.* **66**, 579–601 (1997).
- Diamond, J. M. in *Extinctions* (ed. Nitecki, M. H.) 191–246 (Univ. Chicago Press, Chicago, 1984).
- Pimm, S. L. *The Balance of Nature? Ecological Issues in the Conservation of Species and Communities* (Univ. Chicago Press, Chicago, 1991).
- Caughley, G. Directions in conservation biology. *J. Anim. Ecol.* **63**, 215–244 (1994).
- McKinney, M. L. in *The Biology of Rarity: Causes and Consequences of Rare-Common Differences* (eds Kunin, W. E. & Gaston, K. J.) 110–129 (Chapman & Hall, London, 1997).
- Gaston, K. J. Species-range size distributions: products of speciation, extinction and transformation. *Phil. Trans. R. Soc. Lond. B* **353**, 219–230 (1998).
- Bennett, P. M. & Owens, I. P. F. Variation in extinction risk among birds: chance or evolutionary predisposition? *Proc. R. Soc. Lond. B* **264**, 401–408 (1997).
- Cotgreave, P. & Pagel, M. in *The Biology of Rarity* (eds Kunin, W. E. & Gaston, K. J.) 237–261 (Chapman & Hall, London, 1997).

- Strahan, R. *The Mammals of Australia* (Reed, Sydney, 1995).
- Kirsch, J. A. W., Lapointe, F.-J. & Springer, M. S. DNA-hybridisation studies of marsupials and their implications for metatherian classification. *Aust. J. Zool.* **45**, 211–280 (1997).
- Groves, C. P. & Flannery, T. F. in *Bandicoots and Bilbies* (eds Seebeck, J. H., Wallis, R. L., Brown, P. R. & Kemper, C. M.) 1–11 (Surrey Beatty, Sydney, 1990).
- Close, R. L., Murray, J. D. & Briscoe, D. A. in *Bandicoots and Bilbies* (eds Seebeck, J. H., Brown, P. R., Wallis, R. L. & Kemper, C. M.) 19–27 (Surrey Beatty, Sydney, 1990).
- Baverstock, P. R., Krieg, M. & Birrell, J. Evolutionary relationships of Australian marsupials as assessed by albumin immunology. *Aust. J. Zool.* **37**, 273–287 (1990).
- Archer, M. in *Carnivorous Marsupials* (ed. Archer, M.) 397–443 (R. Zool. Soc. New South Wales, Sydney, 1982).
- Baverstock, P. R., Archer, M., Adams, M. & Richardson, B. J. in *Carnivorous Marsupials* (ed. Archer, M.) 641–650 (R. Zool. Soc. New South Wales, Sydney, 1982).
- Archer, M. Results of the Archbold Expeditions No. 104. Systematics of the marsupials. *Bull. Am. Mus. Nat. Hist.* **168**, 61–224 (1981).
- Flannery, T. F. *Possums of the World* (GEO/Australian Museum, Sydney, 1995).
- Flannery, T. F. in *Kangaroos, Wallabies and Rat-kangaroos* (eds Grigg, G., Jarman, P. & Hume, I.) 1–46 (Surrey Beatty, Sydney, 1989).
- Colgan, D. J. & Flannery, T. F. Biochemical studies of the genus *Petaurus* (Marsupialia: Petauridae). *Aust. J. Zool.* **40**, 245–256 (1992).
- Eldridge, M. D. B., Johnston, P. G. & Close, R. L. Chromosomal rearrangements in rock wallabies, *Petrogale* (Marsupialia: Macropodidae). V. Chromosomal phylogeny of the *lateralis/panicillata* group. *Aust. J. Zool.* **39**, 629–641 (1991).
- Purvis, A. & Rambaut, A. Comparative analysis by independent contrasts (CAIC): an Apple Macintosh application for analysing comparative data. *Comp. Appl. Biosci.* **11**, 247–251 (1995).
- Harvey, P. H. & Pagel, M. D. *The Comparative Method in Evolutionary Biology* (Oxford Univ. Press, Oxford, 1991).
- Pagel, M. D. A method for the analysis of comparative data. *J. Theor. Biol.* **156**, 431–442 (1992).
- May, R. M., Lawton, J. H. & Stork, N. E. in *Extinction Rates* (eds Lawton, J. H. & May, R. M.) 1–24 (Oxford Univ. Press, Oxford, 1995).
- Blackburn, T. M., Lawton, J. H. & Perry, J. N. A method of estimating the slope of upper bounds of plots of body size and abundance in natural animal assemblages. *Oikos* **65**, 107–112 (1992).

Acknowledgements. I thank M. Blows, J. Caley and L. Schwarzkopf for comments; A. Payne for help in compiling data on range size; A. Cockburn for providing facilities at the Division of Botany and Zoology at the Australian National University; and S. Burnett, A. Dennis, M. Evans, A. Horsup, S. Jackson, P. Jarman, L. Leong, R. Martin, A. Melzer, G. Newell, M. Runcie, K. Vernes, P. Winkler, J. Winter and A. Woolnough for unpublished data on population densities. This work was partly supported by a grant from the ARC.

Correspondence and requests for copies of the data set used should be addressed to the author (e-mail: christopher.johnson@jcu.edu.au).

A multimodal language region in the ventral visual pathway

Christian Büchel, Cathy Price & Karl Friston

The Wellcome Department of Cognitive Neurology, Institute of Neurology, Queen Square, London WC1N 3BG, UK

Reading words and naming pictures involves the association of visual stimuli with phonological and semantic knowledge. Damage to a region of the brain in the left basal posterior temporal lobe (BA37), which is strategically situated between the visual cortex and the more anterior temporal cortex, leads to reading and naming deficits^{1,2}. Additional evidence implicating this region in linguistic processing comes from functional neuroimaging studies of reading in normal subjects^{3–7} and subjects with developmental dyslexia^{8,9}. Here we test whether the visual component of reading is essential for activation of BA37 by comparing cortical activations elicited by word processing in congenitally blind, late-blind and sighted subjects using functional neuroimaging. Despite the different modalities used (visual and tactile), all groups of subjects showed a common activation of BA37 by words relative to non-word letter-strings. These findings agree with the proposal that BA37 is an association area that integrates converging inputs from many regions¹⁰. Our study confirms a prediction of theories of brain function that depend on convergence zones; the absence of one input (that is, visual) does not alter the response properties of such a convergence region.

Studies of non-human primates have shown that object and spatial processing are dissociated into interconnected but distinct visual pathways. Visual object processing, including processing of letters and words^{11,12}, is mainly subserved by ventral areas comprising the occipital and inferior temporal cortex. This specialization within the human cortex has been shown by direct cortical recordings¹³, stimulation^{14,15}, functional imaging⁶, and lesion

studies^{2,16}. Ventral regions may also be further specialized for processing faces, coloured objects, and letters^{13,17,18}. Posterior (occipital) regions are activated equally by visual words and letter-strings, whereas more anterior (temporal) regions respond only to words¹³. This has led to the idea that there are successive processing stages within the ventral visual pathway that are concerned with increasingly higher-order characteristics of word or word-like stimuli.

Here we investigated what happens when, because of blindness, language acquisition is restricted to tactile and auditory experience and reading is achieved through the Braille system. Our critical question was: does tactile reading use the same neuronal systems as visual linguistic processing in subjects who have never been exposed to visual words or objects, or does it involve somatosensory association areas (that is, the parietal cortex)? We consider congenital blindness to be a model in which the normal sensory determinants of neurodevelopment in the language system have been perturbed. In a similar sense, late blindness can be seen as a model for a system that is neurodevelopmentally normal but which has subsequently undergone equivalent disruption of normal sensory input relevant to reading.

Early blindness has been shown to lead to crossmodal sensory reorganization, indexed by striate¹⁹ and extrastriate^{19,20} responses to tactile stimuli. We studied higher-order linguistic processing by comparing responses to words and letter-strings within the same modality (tactile or visual). We used a multifactorial positron

emission tomography (PET) design, and involved sighted as well as congenitally and late-blind subjects. Subjects were presented with words or non-word letter-strings while they were engaged in an incidental feature-detection task (see Methods). Half the word conditions contained concrete nouns (for example, 'apple') and the other half contained abstract nouns with low imageability (for example, 'justice'). The stimuli were presented visually to sighted subjects and in the form of Braille to blind subjects.

Activations related to words contrasted with letter-strings were remarkably similar for sighted, late-blind and congenitally blind subjects (Fig. 1), with the greatest activation being expressed in the left posterior basal temporal lobe (BA37) and another activation occurring in the left frontal operculum. A conjunction analysis (see Methods) identified these areas to be significant in all three groups, in the absence of any significant interaction (that is, there were no differential activations among the groups; Fig. 1d)⁵. Concrete and abstract words produced equivalent activation, showing that responses in BA37 are not restricted to words depicting objects. These results (Fig. 1a) replicate those found for sighted subjects in a similar implicit reading task³ and in explicit reading tasks⁴⁻⁷.

Many studies have shown the modality-independent nature of the frontal operculum²¹. Here we focus on the tactile activations of BA37 because this region has been primarily implicated in visual processing^{22,3}. Figure 2 shows the location of the BA37 activation relative to the posterior region (BA19) that responds nonspecifically to visual or tactile stimuli (relative to resting conditions¹⁹ or to an

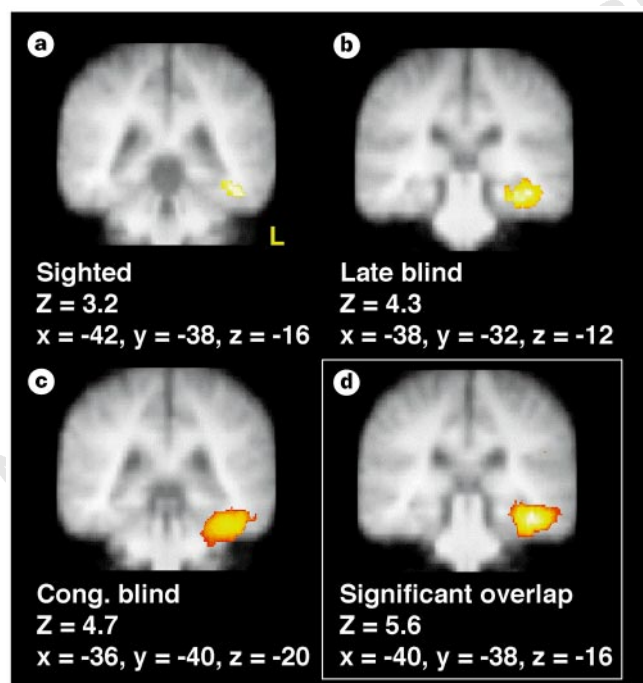


Figure 1 Word-specific activations (responses to words minus responses to consonant letter-strings) in sighted, late-blind and congenitally blind subjects. Words were presented visually for the sighted subjects and using Braille for blind subjects. **a–c**, Significant activations are overlaid on coronal sections of each groups' mean magnetic resonance image. **d**, The activation that is common ($P < 0.05$, corrected for the brain volume analysed) to all three groups according to a conjunction analysis⁵ is shown. Activations in the blind groups (**b, c**) were significant at $P < 0.05$ (corrected). The activation for the sighted group (**a**) was less significant at $P < 0.001$ (uncorrected); we report it here because it was predicted *a priori*^{2,6,7,13,17}. All maps have a threshold of $P < 0.01$ (uncorrected) for visualization.

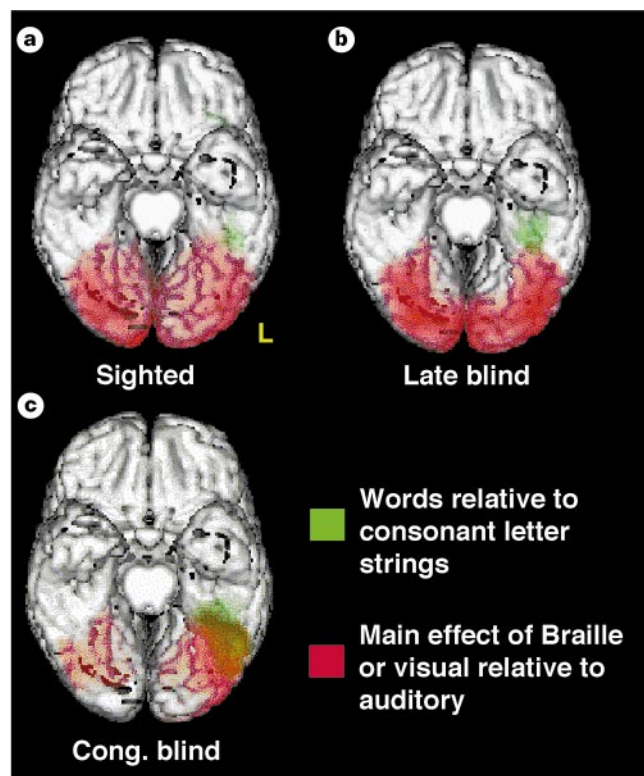


Figure 2 Demonstration of the anterior/posterior dissociation between word-specific and nonspecific cortical activations on ventral views of the brain surface (excluding the cerebellum). Nonspecific activations (that is, responses to word and non-word letter-strings minus responses to word and non-word auditory stimuli) are shown in red on the cortical surface. Word-specific activations (that is, responses to visual or tactile words minus responses to non-word letter-strings) are anterior and lateral, as shown in green. All maps have a threshold of $P < 0.01$ (uncorrected) for visualization to show the extent of the activations. The intensity of colours reflects depth from the cortical surface (bright for superficial activations). **a**, Sighted group; **b**, late-blind group; **c**, congenitally blind group.

auditory baseline²⁰). The more anterior, BA37 response (Fig. 2) is word-specific. This anterior/posterior dissociation for word-specific (anterior) and nonspecific (posterior) responses along the fusiform gyrus confirms findings from direct cortical recordings¹³. However, the anterior area, activated by all subject groups represents only a small portion within BA37, illustrating the functional diversity of this association area.

Late-blind and sighted subjects showed nearly identical activation patterns (Fig. 1a, b). This similarity may reflect the fact that the late-blind subjects developed visual reading skills before the onset of blindness, and is consistent with the existence of an established visual reading system that is subsequently recruited for tactile reading. More surprising was the fact that we found activations in the posterior basal temporal lobe of congenitally blind subjects (Fig. 1c), in whom this cortical system had never been engaged by visual processing. Although human studies have emphasized the visual properties of BA37 (refs 22, 23), our findings show that tactile information is also processed in BA37. This multimodality concurs with anatomical data about connectivity in precursor regions of BA37 in the monkey; these data indicate convergence of auditory, visual and tactile projections²⁴.

The predominant deficit after lesions of BA37 in humans are alexia and (visual) anomia, but an associated tactile anomia has also been reported^{1,25}. The tactile deficit in sighted subjects could be due either to the fact that tactile information is intermediately transformed into a visual representation²³, which BA37 then integrates with lexical information, or to the fact that BA37 directly integrates tactile information with lexical information. Our data from congenitally blind subjects show that BA37 has multimodal properties that integrate tactile inputs with linguistic knowledge even in the absence of tactile–visual associations.

We can discount the possibility that the common basal temporal lobe activations we found represent activations due to visual or tactile imagery²³ of objects depicted by the words, because half of the words presented were abstract (for example, ‘justice’) and these words produced the same posterior temporal responses as concrete words (with no significant difference in activation). This reasoning is extended by the absence of any possible visual imagery and by restrictions on tactile imagery (that is, object size) in congenitally blind subjects.

In sighted children, language develops by the association of a spoken word with visual information (for example, the picture of a ball or later the visual word ‘ball’). Conversely, in congenitally blind children it is the association of tactile input with the spoken word. Therefore, although sighted and blind subjects have auditory experience in common, they differ in the sensory input to which they attach this phonological/auditory information. Our data show that the tactile and auditory experience available to blind subjects is sufficient for BA37 to integrate sensory information with linguistic knowledge.

Complementary evidence to support the hypotheses that parts of BA37 are involved in linguistic processing comes from studies of developmental dyslexia. In a neuroimaging study using the same task paradigm, we showed that developmental dyslexics showed reduced activation, compared with controls, in BA37 (ref. 8), consistent with earlier functional imaging studies⁹. The adult developmental dyslexic subjects had profound difficulties in learning to read despite (above) average intelligence, normal educational opportunity, and absence of sensory impairment. Nevertheless, their reading ability was well-compensated and they were able to read the words as accurately and within the same time constraints as normal control subjects. The reduced responses to words in BA37 in these subjects shows the enduring effect of their developmental disorder.

The activation of BA37 during reading³ and picture naming⁵ indicates BA37 may link internal knowledge with arbitrary symbols. Is this link linguistic or prelinguistic? A patient with a small lesion in

BA37 (ref. 1) showed an isolated ‘phonological anomia’ with alexia, suggesting a linguistic role for BA37. Patients with anomia and alexia may fail to retrieve the name associated with an object or read a word. At the same time, they can identify presented objects by detailed verbal description and read words after naming letters sequentially²⁶, indicating intact conceptual knowledge²⁷. The evidence that BA37 links converging inputs suggest that this area may not contain records of linguistic codes *per se*. BA37 may be an area that promotes activity in other distributed sites that jointly give rise to lexical or conceptual access (see also ref. 10). Such a region should be resistant to major variations in sensory input characteristics¹⁰. This prediction is confirmed by our data, which show that BA37 is activated when blind subjects read Braille. □

Methods

PET scanning and data analysis. PET scans were acquired with an ECAT EXACT HR+ scanning system (CTI Siemens) in three-dimensional mode. An antecubital vein cannula was used to administer 350 mBq of H₂¹⁵O. We acquired the data in one 90-s frame with about 8 min between scans. Images were reconstructed by filtered back projection into a 128 × 128 × 63 image matrix (voxel size 2.1 × 2.1 × 2.4 mm) using measured attenuation correction. The data were analysed with SPM96 (Wellcome Dept Cognitive Neurology; <http://www.fil.ion.ucl.ac.uk/spm>). Images were realigned to correct for interscan movements, co-registered with the subject’s structural magnetic resonance image to allow overlaying of functional data on the subjects’ structural data, and spatially normalized into the space defined by the atlas of J. Talairach and P. Tournoux to allow group analyses. The data were smoothed with a 16-mm gaussian filter to account for residual intersubject differences. Statistical parametric maps were derived with the appropriate contrasts of condition-specific effects. The contrasts of interest in this study tested for the difference between reading words and non-words for all three subject groups, thereby revealing voxels in which the regional cerebral bloodflow was significantly higher during word reading compared with during non-word reading. These voxel-based *t*-statistics are transformed to *Z*-statistics and assembled in a statistical parametric map, SPM{*Z*}, that can be seen as an image of *Z*-statistics for a single comparison. Statistical inference on the SPM{*Z*} was corrected for multiple comparisons using the theory of gaussian fields²⁸.

We report only activations reflecting differences between word and non-word processing that were common to blind and sighted subjects. Significant interactions (that is, differences between groups) were found but are not reported here. Our analysis was voxel-based and therefore not biased towards any specific region of the brain.

Subjects and stimuli. We studied six sighted subjects (mean age 26.8 ± 4 years), six congenitally blind, proficient Braille readers (mean age 49.2 ± 12 years), and three late-blind subjects (mean age of onset of blindness 18.3 ± 3.8 years; mean age 45 ± 7.6 years). Eight out of nine blind subjects read with their right hand. One congenitally blind reader used both hands. The causes of blindness in the congenital group were retinopathy of prematurity (2 of 6), optic nerve abnormalities (3 of 6) and congenital glaucoma (1 of 6). All late-blind subjects experienced progressive visual impairment before complete blindness. The initial pathology in the late-blind group comprised iridocyclitis, trauma and microphthalmia with ensuing calcifying cataracts. The late-blind subjects could read print before losing their sight and were matched in terms of Braille reading to the congenitally blind group (finding subjects that meet these criteria is difficult, hence the small number of late-blind subjects). All blind subjects had no light perception. All subjects were right-handed; one subject was female. The study design was multifactorial, with three factors being altered: modality (visual, tactile or aural), stimulus (word or non-word) and imageability (high or low) within the word level of the previous factor.

Each subject underwent 12 scans, during which we presented either words or non-words at a rate of one per 2.5 s. For the sighted group, visual stimuli were presented on a computer monitor (5 degrees vertically with variable word length horizontally) for 1 s. For the blind groups, six dot, non-contracted (one Braille character per letter) Braille strings were presented on a PC-driven electronic Braille display (ALVA, Netherlands) for 2.3 s. Non-words were strings of consonants (for example, CDXSD). We chose to compare consonant letter-strings to words instead of pseudowords (such as foop) because the face validity

of this comparison is easier to establish, especially for Braille readers about whom there is little information on pseudoword processing.

Auditory stimuli had the same characteristics as the visual/tactile stimuli except that the non-words were reversed words. During half of the word-presentation conditions, we presented abstract words (low imageability 250–400 (according to the Medical Research Council psycholinguistic database)) and concrete words (high imageability 550–700). The task in all conditions was a feature-detection task involving incidental word processing as described previously³. Targets were ascenders in the visual conditions (for example in the letters l, d and t), one or more raised dot number 6 (the right lower dot) in the Braille words, and a short tone in the auditory conditions. The use of implicit processing allowed us to match attentional set between conditions and to monitor performance. Words and non-words were matched for the number of characters (5.6 ± 1.3), syllables (1.7 ± 0.5) and occurrences of targets (49%). Responses were made by pressing a button with the thumb of the right hand. The instructions were: "Read/listen to the word and if you find/hear a target, press the button". Target detection exceeded 85% in all groups. The mean reaction time for visual/tactile presentation was 552 ± 116 ms for the sighted group, $1,605 \pm 281$ ms for the congenitally blind group and $1,809 \pm 364$ ms for the late-blind group.

Received 6 April; accepted 27 May 1998.

1. Rapsak, S. Z., Gonzalez Rothi, L. J. & Heilman, K. M. Phonological alexia with optic and tactile anomia: a neuropsychological and anatomical study. *Brain Lang.* **31**, 109–121 (1987).
2. Damasio, A. R. & Damasio, H. The anatomical basis of pure alexia. *Neurology* **33**, 1573–1583 (1983).
3. Price, C. J., Wise, R. J. S. & Frackowiak, R. S. J. Demonstrating the implicit processing of visually presented words and pseudowords. *Cerebral Cortex* **6**, 62–70 (1996).
4. Petersen, S. E., Fox, P. T., Snyder, A. Z. & Raichle, M. E. Activation of extrastriate and frontal cortical areas by words and word-like stimuli. *Science* **249**, 1041–1044 (1990).
5. Price, C. J. & Friston, K. J. Cognitive conjunction: a new approach to brain activation experiments. *Neuroimage* **5**, 261–270 (1997).
6. Puce, A., Allison, T., Asgari, M., Gore, J. C. & McCarthy, G. Differential sensitivity of human visual cortex to faces, letterstrings, and textures: a functional magnetic resonance imaging study. *J. Neurosci.* **16**, 5205–5215 (1996).
7. Bookheimer, S. Y., Zeffiro, T. A., Blaxton, T., Gaillard, W. & Theodore, W. Regional cerebral blood flow during object naming and word reading. *Hum. Brain Mapp.* **3**, 93–106 (1995).
8. Brunswick, N., McCloy, E., Price, C., Frith, C. D. & Frith, U. Explicit and implicit processing of words and pseudowords by adult developmental dyslexics: a search for Wernicke's Wortschatz? *Brain* (submitted).
9. Rumsey, J. M. et al. A positron emission tomographic study of impaired word recognition and phonological processing in dyslexic men. *Arch. Neurol.* **54**, 562–573 (1997).
10. Damasio, A. R. & Damasio, H. In *Large Scale Neuronal Theories of the Brain* (eds Koch, C. & Davis, J. L.) 61–74 (MIT Press, Cambridge, MA, 1994).
11. Ungerleider, L. G. & Haxby, J. V. 'What' and 'where' in the human brain. *Curr. Opin. Neurobiol.* **4**, 157–165 (1994).
12. Haxby, J. V. et al. Dissociation of object and spatial visual processing pathways in human extrastriate cortex. *Proc. Natl Acad. Sci. USA* **88**, 1621–1625 (1991).
13. Nobre, A. C., Allison, T. & McCarthy, G. Word recognition in the human inferior temporal lobe. *Nature* **372**, 260–263 (1994).
14. Luders, H. et al. Basal temporal language area demonstrated by electrical stimulation. *Neurology* **36**, 505–510 (1986).
15. Burnstine, T. H. et al. Characterization of the basal temporal language area in patients with left temporal lobe epilepsy. *Neurology* **40**, 966–970 (1990).
16. Farah, M. J. *Visual Agnosia* (MIT Press, Cambridge, MA, 1990).
17. Allison, T., McCarthy, G., Nobre, A., Puce, A. & Belger, A. Human extrastriate visual cortex and the perception of faces, words, numbers, and colors. *Cerebral Cortex* **4**, 544–554 (1994).
18. Kanwisher, N., McDermott, J. & Chun, M. M. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J. Neurosci.* **17**, 4302–4311 (1997).
19. Sadato, N. et al. Activation of the primary visual cortex by Braille reading in blind subjects. *Nature* **380**, 526–528 (1996).
20. Büchel, C., Price, C. J., Frackowiak, R. S. J. & Friston, K. J. Different activation patterns in the occipital cortex of late and congenitally blind subjects. *Brain* **121**, 409–419 (1998).
21. Fiez, J. A. et al. PET studies of auditory and phonological processing—effects of stimulus characteristics and task demands. *J. Cog. Neurosci.* **7**, 357–375 (1995).
22. Martin, A., Haxby, J. V., Lalonde, F. M., Wiggs, C. L. & Ungerleider, L. G. Discrete cortical regions associated with knowledge of color and knowledge of action. *Science* **270**, 102–105 (1995).
23. Desposito, M. et al. A functional MRI study of mental image generation. *Neuropsychol.* **35**, 725–730 (1997).
24. Suzuki, W. A., Zola-Morgan, S., Squire, L. R. & Amaral, D. G. Lesions of the perirhinal and parahippocampal cortices in the monkey produce long-lasting memory impairment in the visual and tactile modalities. *J. Neurosci.* **13**, 2430–2451 (1993).
25. De Renzi, E., Zambolin, A. & Crisi, G. The pattern of neuropsychological impairment associated with left posterior cerebral-artery infarcts. *Brain* **110**, 1099–1116 (1987).
26. Friedman, R. F., Ween, J. E. & Albert, M. L. In *Clinical Neuropsychology* (eds Heilman, K. M. & Valenstein, E.) 37–62 (Oxford Univ. Press, Oxford, 1993).
27. Foundas, A. L., Daniels, S. K. & Vasterling, J. J. Anomia: case studies with lesion localization. *Neurocase* **4**, 35–43 (1998).
28. Friston, K. J. et al. Statistical parametric maps in functional imaging: a general linear approach. *Hum. Brain Mapp.* **2**, 189–210 (1995).

Acknowledgements. The authors are supported by the Wellcome Trust. We thank A. Brennan, A. Carrol, G. Lewington, J. Galliers and S. Grootenk for their help with data collection, R. Frackowiak, J. Coull and A. Kleinschmidt for an internal review of this manuscript, and M. Rötger for discussions. We are indebted to our subjects, especially the blind subjects and the Royal National Institute for the Blind in the UK.

Correspondence and requests for materials should be addressed to C.B. (e-mail: c.buechel@fil.ion.ucl.ac.uk).

Control of pain initiation by endogenous cannabinoids

Antonio Calignano, Giovanna La Rana, Andrea Giuffrida* & Daniele Piomelli*

Dipartimento di Farmacologia Sperimentale, Università di Napoli, 49 via D. Montesano, Napoli 80131, Italy

* The Neurosciences Institute, 10640 J. J. Hopkins Drive, San Diego, California 92121, USA

The potent analgesic effects of cannabis-like drugs^{1–4} and the presence of CB1-type cannabinoid receptors in pain-processing areas of the brain and spinal cord^{5,6} indicate that endogenous cannabinoids such as anandamide⁷ may contribute to the control of pain transmission within the central nervous system (CNS)⁸. Here we show that anandamide attenuates the pain behaviour produced by chemical damage to cutaneous tissue by interacting with CB1-like cannabinoid receptors located outside the CNS. Palmitylethanolamide (PEA), which is released together with anandamide from a common phospholipid precursor⁹, exerts a similar effect by activating peripheral CB2-like receptors. When administered together, the two compounds act synergistically, reducing pain responses 100-fold more potently than does each compound alone. Gas-chromatography/mass-spectrometry measurements indicate that the levels of anandamide and PEA in the skin are enough to cause a tonic activation of local cannabinoid receptors. In agreement with this possibility, the CB1 antagonist SR141716A and the CB2 antagonist SR144528 prolong and enhance the pain behaviour produced by tissue damage. These results indicate that peripheral CB1-like and CB2-like receptors participate in the intrinsic control of pain initiation and that locally generated anandamide and PEA may mediate this effect.

Although pain perception is thought to be controlled mainly by neurotransmitter systems that operate within the CNS¹⁰, antinociceptive (pain-relieving or -preventing) mechanisms also occur in peripheral tissues. For example, endogenous opioid peptides released from activated immune cells during inflammation inhibit pain transmission by interacting with opioid receptors on peripheral sensory nerve endings¹¹. To determine whether endogenous cannabinoids have an analogous function to that of opioids, we used the formalin test, a behavioural model of injury-induced acute and tonic pain¹². Injection of dilute formalin into the hind paws of freely moving rodents evokes a pain behaviour consisting of two temporally distinct phases of licking and flexing of the injected limb¹². An early phase involving acute activation of pain-sensing C fibres¹³ begins immediately after formalin administration, reaches a peak within 5 min, and then rapidly declines. After an interval of 10–15 min, a second phase of sustained pain behaviour appears, in which sensory fibre activity is accompanied by inflammation¹⁴ and central sensitization¹⁵.

Table 1 Effects of anandamide and PEA on the behavioural response to acute thermal stimuli

Time (min)	Hot-plate latencies (s)			
	Vehicle	AEA	PEA	AEA/PEA
5	24 ± 2	24 ± 4	22 ± 3	25 ± 2
10	24 ± 2	23 ± 3	23 ± 3	26 ± 4
15	22 ± 3	29 ± 3	21 ± 2	30 ± 3
20	25 ± 2	35 ± 2*	25 ± 4	37 ± 3*
30	25 ± 3	40 ± 3*	26 ± 5	39 ± 2*
40	25 ± 3	30 ± 3	22 ± 4	35 ± 5
60	26 ± 2	28 ± 3	24 ± 2	26 ± 3

Time delays to the response were measured at certain times (in min) after intracerebroventricular administration of vehicle (10% DMSO in saline, 5 µl), anandamide (AEA, 10 µg), PEA (10 µg) or anandamide plus PEA (10 µg each). Asterisk indicates $P < 0.01$ ($n = 6$).

In mice, the early phase of pain behaviour was blocked when anandamide was injected into the paw together with formalin, whereas both phases were blocked by the synthetic cannabinoid agonists WIN-55212-2 and HU-210 (Fig. 1a; data not shown). These analgesic effects were prevented by systemic administration of the CB1 antagonist SR141716A (Fig. 1a), but not of the CB2 antagonist SR144528 (Fig. 1a)¹⁶ or of the opioid antagonist naloxone (0.2 mg per kg, intravenous; data not shown). The lack of effect of anandamide on late-phase pain behaviour may be explained by the short lifespan of this compound, which undergoes rapid biological inactivation in tissues^{17,18}. In support of this suggestion, the inactivation-resistant analogue methanandamide inhibited pain behaviour during the entire test (Fig. 1a).

Local anandamide injections were not accompanied by central signs of cannabimimetic activity, indicating a peripheral site of action. To test this possibility, we measured the antinociceptive potency of anandamide following local (intraplantar, i.pl.), intravenous (i.v.) or intraperitoneal (i.p.) administrations. Anandamide was 100 times more potent in preventing formalin-evoked pain behaviour when injected locally rather than intravenously, with half-maximal inhibitory doses (ID₅₀ values) of 0.1 mg per kg and 10 mg per kg, respectively (Fig. 1b). Anandamide had no significant effect when injected intraperitoneally (Fig. 1b). As a further test, we determined the biodistribution of ³H-labelled anandamide 10 min after i.pl. injection in rats. In three experiments, we found that 94% of recovered ³H-labelled anandamide remained associated with the injected paw (6648 ± 820 d.p.m. per g, mean ± s.e.m.), whereas little or no radioactivity above background was detected in fore-brain, cerebellum and spinal cord (79 ± 19; 165 ± 67; and 90 ± 48 d.p.m. per g, respectively). These results indicate that anan-

damide inhibits nociception after formalin injection by activating CB1-like receptors, which may be located on peripheral endings of sensory neurons involved in pain transmission^{3,6}.

Anandamide is the ethanolamide of arachidonic acid and is thought to be produced by phosphodiesterase-mediated cleavage of *N*-arachidonyl phosphatidylethanolamine^{9,19}. Enzymatic cleavage of other *N*-acyl phosphatidylethanolamines may give rise to additional fatty acylethanolamides, the physiological roles of which are still poorly understood²⁰. PEA, an acylethanolamide found in neural and non-neural tissues, inhibits mast-cell activation and reduces inflammatory responses^{21,22} by a mechanism that may involve binding to CB2-like receptors²³. The molecular identity of these receptors is unknown, although they are likely to be distinct from the CB2 receptors whose encoding genes have been cloned, as PEA shows little or no affinity for these receptors²⁴. We found that PEA, but not two closely related analogues, inhibited both early and late phases of formalin-evoked pain behaviour after i.pl. injection in mice (Fig. 2a). This effect may not be explained by the anti-inflammatory properties of PEA; 30 min after injection, the analgesic effects of PEA were not accompanied by a reduction in inflammatory oedema (volumes injected into the paw in ml were: control, 0.18 ± 0.003; formalin, 0.37 ± 0.006; formalin plus 50 µg PEA, 0.35 ± 0.006), which became apparent only 1 h after formalin administration²².

The analgesia produced by PEA was reversed by administration of the CB2 antagonist SR144528 (Fig. 2a), whereas the CB1 antagonist SR141716A and the opioid antagonist naloxone were ineffective (Fig. 2a; data not shown). In addition, PEA was more potent when administered locally (i.pl.) than systemically (i.v. or i.p.) (Fig. 2b). Together, these results indicate that PEA exerts antinociceptive

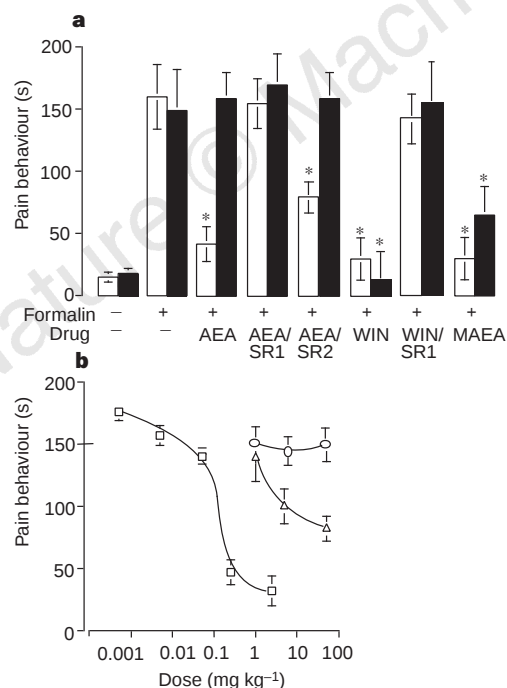


Figure 1 Anandamide inhibits formalin-evoked nociception by activating peripheral CB1-like cannabinoid receptors. **a**, Effects of anandamide (AEA, 50 µg i.pl.), WIN-55212-2 (WIN, 500 µg i.pl.) and methanandamide (MAEA, 50 µg i.pl.), in the absence or presence of the CB1 antagonist SR141716A (SR1, 0.1 mg per kg i.v.) or the CB2 antagonist SR144528 (SR2, 0.1 mg per kg i.v.). Open columns represent the early phase and filled columns the late phase of formalin-evoked nociception. Asterisk indicates $P < 0.01$ ($n = 12-18$ for each condition). **b**, Dose-dependent antinociceptive effects of anandamide following i.pl. (squares), i.v. (triangles) or i.p. (circles) administration.

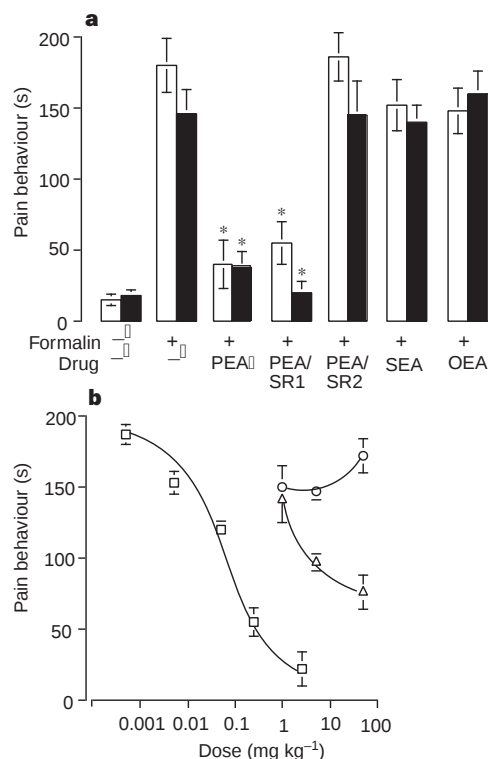


Figure 2 PEA inhibits formalin-evoked nociception by activating peripheral CB2-like cannabinoid receptors. **a**, Effects of PEA (50 µg i.pl.), stearyl ethanolamide (SEA, 50 µg i.pl.) and oleyl ethanolamide (OEA, 50 µg i.pl.), in the absence or presence of the CB1 antagonist SR141716A (SR1, 0.1 mg per kg i.v.) or the CB2 antagonist SR144528 (SR2, 0.1 mg per kg i.v.). Open columns represent the early phase and filled columns the late phase of formalin-evoked nociception. Asterisk indicates $P < 0.01$ ($n = 12-18$). **b**, Dose-dependent antinociceptive effects of PEA following i.pl. (squares), i.v. (triangles) or i.p. (circles) administration.

effects that are mediated by peripheral CB2-like receptors. The cellular localization of such receptors and their possible structural relationship with the CB2 receptor whose gene has been cloned, which is primarily expressed in immune cells²⁵, are unknown.

The fact that anandamide and PEA activate pharmacologically distinct receptors and that these two substances can be produced simultaneously in tissues⁹ prompted us to examine their possible interactions *in vivo*. When injected together in equal amounts, anandamide and PEA inhibited the early phase of formalin-evoked pain behaviour with a potency that was approximately 100-fold greater than each of the compounds separately (Fig. 3a). A similar synergistic potentiation occurred in the late phase, on which anandamide had no effect when given alone (Figs 1a and 3b). Earlier administration of either CB1 or CB2 antagonists entirely blocked the response (Fig. 3c). This interaction did not appear to involve pain-processing structures within the brain: injection of PEA in the cerebral ventricles did not affect the behavioural

responses to acute thermal stimuli, assessed in the hot-plate test, and did not enhance the inhibitory activity of anandamide administered by the same route (Table 1). These results indicate that the parallel activation of peripheral CB1- and CB2-like receptors by anandamide and PEA results in a synergistic inhibition of peripheral pain initiation.

To test this idea further, we determined the intrinsic effects of CB1 and CB2 antagonists on formalin-evoked pain behaviour (Fig. 4). Blockade of CB1 receptors with SR141716A produced significant hyperalgesia (Fig. 4a)⁸. This effect was particularly pronounced after local injection of the drug, which resulted in a 10-min prolongation of the early nociceptive phase and in a two- to threefold increase in pain behaviour during the entire testing period (Fig. 4b). In contrast, systemic administration of the CB2 antagonist SR144528 caused a selective enhancement of early-phase, but not of late-phase, responses (Fig. 4a). The selectivity of this effect may not result from a rapid elimination of SR144528 after early phase, as the drug reversed PEA-induced antinociception during both early- and late-phase pain behaviour (Fig. 2). SR144528 could not be administered locally because of its limited solubility in the injection vehicle.

Although the hyperalgesic effects of CB1 and CB2 antagonists may be accounted for by their inverse agonist properties^{16,26}, two lines of evidence suggest that these drugs acted by removing an endogenous cannabinoid tone. First, gas-chromatography/mass-spectrometry analyses showed that anandamide and PEA are present in

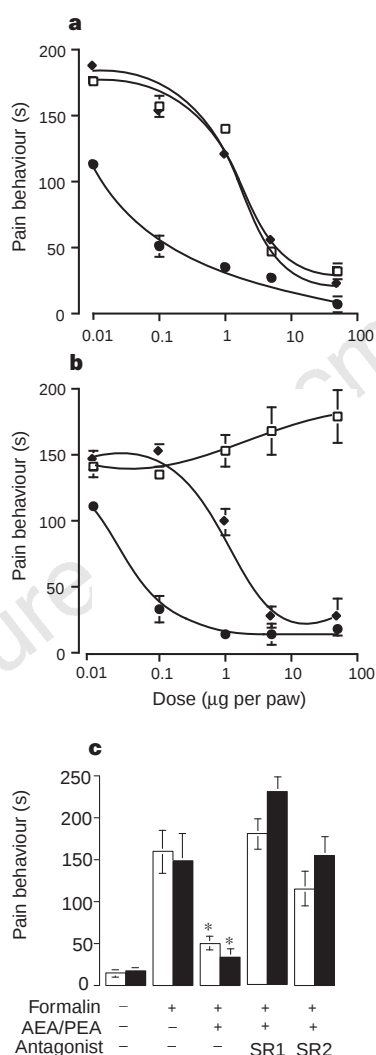


Figure 3 Anandamide and PEA synergistically inhibit formalin-evoked nociception. **a**, Early phase. **b**, Late phase (open squares, anandamide; filled diamonds, PEA; filled circles, anandamide plus PEA). Equal amounts of anandamide and PEA were administered by i.pl. injection at the doses indicated on the abscissa. **c**, The CB1 antagonist SR141716A (SR1) and the CB2 antagonist SR144528 (SR2) prevent the effects of anandamide plus PEA (0.1 µg each of anandamide and PEA). Open columns represent the early phase and closed columns the late phase of formalin-evoked nociception. Antagonists were administered by i.v. injection at the dose of 0.1 mg per kg. Asterisk denotes $P < 0.01$ ($n = 12$).

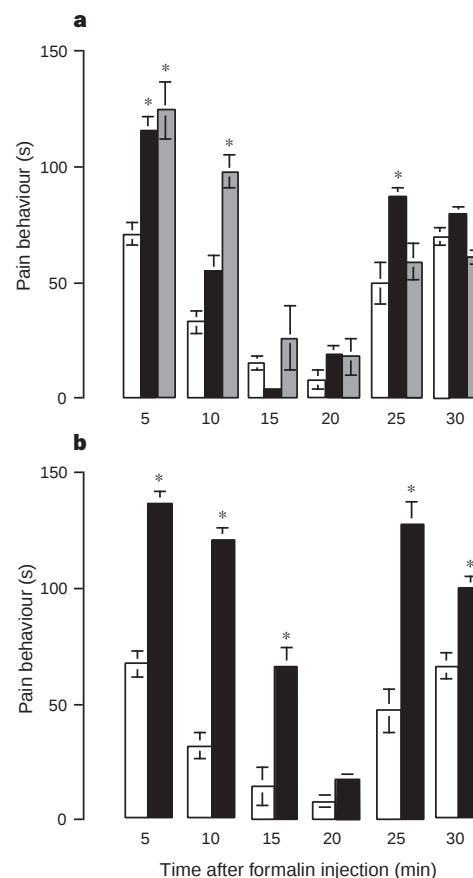


Figure 4 Intrinsic hyperalgesic effects of cannabinoid antagonists on the time course of formalin-evoked nociception. **a**, Effects of systemic administration of the CB1 antagonist SR141716A (filled bars) and the CB2 antagonist SR144528 (grey shaded bars). The drugs were administered by i.v. injection at 0.1 mg per kg. The responses to formalin alone are shown by open bars. **b**, Effects of local administration of SR141716A (filled bars). 10 µg of SR141716A were administered with formalin by i.pl. injection. SR144528 was not soluble in 10% DMSO, and could not be injected locally. Asterisk denotes $P < 0.01$ ($n = 6$).

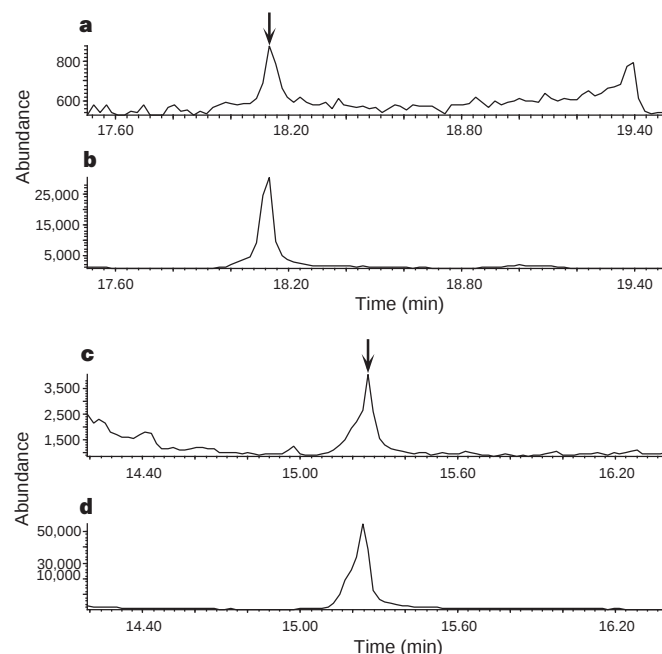


Figure 5 Identification by gas chromatography/mass spectrometry of anandamide and PEA in rat paw skin. **a, b**, Representative tracings for selected fragments characteristic of endogenous anandamide (**a**, mass-to-charge ratio $m/z = 404$, $[M - 15]^+$, a fragment produced by loss of one methyl group) and synthetic $[^2H_4]$ anandamide (**b**, $m/z = 408$) which was added to the samples as an internal standard. **c, d**, Representative tracings for fragments characteristic of endogenous PEA (**c**, $m/z = 356$, $[M - 15]^+$) and standard $[^2H_4]$ PEA (**d**, $m/z = 360$). The results are from one experiment and are typical of eight experiments.

rat paw skin (Fig. 5). By comparison with internal deuterated standards, we measured 49 ± 9 pmol of anandamide and 692 ± 119 pmol of PEA per g of tissue ($n = 8$). These amounts are five- to tenfold higher than those measured by the same method in rat brain and plasma^{27,28}, and are probably sufficient to activate cannabinoid receptors^{7,23}. Second, the CB2 antagonist SR144528 enhanced nociception selectively during the early phase of the formalin response (Fig. 4a), a result inconsistent with an inverse agonist effect. Thus, a parsimonious interpretation of our findings is that endogenous PEA acting at CB2-like receptors may participate in attenuating the early stages of nociception, whereas endogenous anandamide acting at CB1-like receptors may have a more sustained modulatory effect on both acute and tonic pain.

Our results show that the tonic activation of local CB1-like and CB2-like receptors may regulate pain initiation in cutaneous tissue. These findings support the possibility that endogenous cannabinoids, in addition to their spinal and supraspinal sites of action⁸, may participate in buffering emerging pain signals at sites of tissue injury. The results also indicate that selective agonists for the CB2-like receptor activated by PEA, or peripherally administered CB1-like/CB2-like agonists, may reduce pain without the dysphoric side effects and perceived abuse potential typical of centrally acting cannabinimimetic or opiate drugs.

Note added in proof: After submission of this paper, Jaggar *et al.*²⁹ have observed an analgesic effect of systemically administered PEA in formalin-induced pain.

Methods

Drugs. Anandamide, PEA, stearyl ethanolamide and oleyl ethanolamide were synthesized following standard procedures⁷; SR144528 $[N-[(1S)\text{-endo-1, 3, 3-trimethyl bicyclo [2.2.1] heptan-2-yl]-5-(4-chloro-3-methylphenyl)-1-(4-methylbenzyl)-pyrazole-3-carboxamide}]$ was a gift from Sanofi; SR141716A $\{[N-(\text{piperidin-1-yl})-5-(4-chlorophenyl)-1-(2,4-dichlorophenyl)-4-methyl-1-H-pyrazole-3-carboxamide-HCl]\}$ was provided by RBI as part of the Chemical

Synthesis Program of the NIMH; all other drugs were from Tocris. The drugs were dissolved in dimethylsulphoxide (DMSO) and administered in physiological saline containing 10% DMSO; volumes injected were 0.2 ml i.v. and i.p. and 0.01 ml i.p.l.

Nociceptive tests. Saline (10 μ l) containing 5% formalin and 10% DMSO was injected subcutaneously into the hind paws of male Swiss mice (20–25 g in weight; Nossan). The duration of paw licking was monitored by an observer blind to the experimental treatment for periods of 0–15 min (early phase) and 15–30 min (late phase) after formalin administration. Initial tests determined that the presence of 10% DMSO in the injection vehicle did not significantly affect formalin responses (data not shown). For the experiment in Fig. 4, pain behaviour was monitored for 5-min periods during the 30 min following formalin injection. Time delays to escape jumping or hind-paw licking were measured on a plate heated to 55.5 °C, according to standard procedures¹⁸.

Inflammation. Inflammatory oedemas were produced in the hind paws of Swiss mice by injection of saline (10 μ l) containing 5% formalin and 10% DMSO, and were measured with a plethysmometer²² (Ugo Basile).

Gas chromatography/mass spectrometry (GC/MS). Hairless paw skin tissue was excised from Wistar rats anaesthetized with nembutal, and immediately immersed in cold methanol to prevent alterations in tissue lipids. After homogenization and chloroform/methanol extraction, high-performance liquid chromatography and quantitative analysis of the trimethyl silyl derivatives of anandamide and PEA by isotope dilution GC/MS were carried out as described²⁸.

Data analysis. Results are expressed as means $\pm$ s.e.m. The significance of differences between groups was evaluated using analysis of variance with a subsequent Dunnett's test.

Received 19 January; accepted 20 April 1998.

- Martin, B. R., Balster, R. L., Razdan, R. K., Harris, L. S. & Dewey, W. L. Behavioral comparisons of the stereoisomers of tetrahydrocannabinols. *Life Sci.* **20**, 565–574 (1981).
- Martin, W. J., Lai, N. K., Patrick, S. L., Tsou, K. & Walker, J. M. Antinociceptive actions of cannabinoids following intraventricular administration in rats. *Brain Res.* **629**, 300–304 (1993).
- Tsou, K. *et al.* Suppression of noxious stimulus-evoked expression of FOS protein-like immunoreactivity in rat spinal cord by a selective cannabinoid agonist. *Neuroscience* **70**, 791–798 (1996).
- Lichtman, A. H., Cook, S. A. & Martin, B. R. Investigation of brain sites mediating cannabinoid-induced antinociception in rats: evidence supporting periaqueductal gray involvement. *J. Pharmacol. Exp. Ther.* **276**, 585–593 (1996).
- Herkenham, M. *et al.* Cannabinoid receptor localization in brain. *Proc. Natl Acad. Sci. USA* **87**, 1932–1936 (1990).
- Tsou, K., Brown, S., Sañudo-Peña, M. C., Mackie, K. & Walker, J. M. Immunohistochemical distribution of cannabinoid CB1 receptors in the rat central nervous system. *Neuroscience* **83**, 393–411 (1998).
- Devane, W. *et al.* Isolation and structure of a brain constituent that binds to the cannabinoid receptor. *Science* **258**, 1946–1949 (1992).
- Richardson, J. D., Aanonsen, L. & Hargreaves, K. M. Hypoactivity of the spinal cannabinoid system results in NMDA-dependent hyperalgesia. *J. Neurosci.* **18**, 451–457 (1998).
- Di Marzo, V. *et al.* Formation and inactivation of endogenous cannabinoid anandamide in central neurons. *Nature* **372**, 686–691 (1994).
- Fields, H. L. *Pain* (McGraw-Hill, New York, 1987).
- Stein, C. The control of pain in peripheral tissue by opioids. *New Engl. J. Med.* **332**, 1685–1690 (1995).
- Dubuisson, D. & Dennis, S. G. The formalin test: a quantitative study of the analgesic effects of morphine, meperidine, and brain stem stimulation in rats and cats. *Pain* **4**, 161–174 (1977).
- Dickenson, A. H. & Sullivan, A. F. Subcutaneous formalin-induced activity of dorsal horn neurones in the rat: differential response to an intrathecal opiate administered pre or post formalin. *Pain* **30**, 349–360 (1987).
- Rosland, J. H., Tjølsen, A., Maehle, B. & Hole, K. The formalin test in mice: effect of formalin concentration. *Pain* **42**, 235–242 (1990).
- Coderre, T. J. & Melzack, R. The contribution of excitatory amino acids to central sensitization and persistent nociception after formalin-induced tissue injury. *J. Neurosci.* **12**, 3665–3670 (1992).
- Rinaldi-Carmona, M. *et al.* SR 144528, the first potent and selective antagonist of the CB2 cannabinoid receptor. *J. Pharmacol. Exp. Ther.* **284**, 644–650 (1998).
- Cravatt, B. F. *et al.* Molecular characterization of an enzyme that degrades neuromodulatory fatty-acid amides. *Nature* **384**, 83–87 (1996).
- Beltramo, M. *et al.* Functional role of high-affinity anandamide transport, as revealed by selective inhibition. *Science* **277**, 1094–1097 (1997).
- Cadas, H., di Tomaso, E. & Piomelli, D. Occurrence and biosynthesis of endogenous cannabinoid precursor, N-arachidonoyl phosphatidylethanolamine, in rat brain. *J. Neurosci.* **17**, 1226–1242 (1997).
- Schmid, H. H. O., Schmid, P. C. & Natarajan, V. The N-acylation-phosphodiesterase pathway and cell signalling. *Chem. Phys. Lipids* **80**, 133–142 (1996).
- Aloe, L., Leon, A. & Levi-Montalcini, R. A proposed autacoid mechanism controlling mastocyte behaviour. *Agents Actions* **39**, C145–C147 (1993).
- Mazzari, S., Canella, R., Petrelli, L., Marcolongo, G. & Leon, A. N-(2-hydroxyethyl)hexadecanamide is orally active in reducing edema formation and inflammatory hyperalgesia by down-modulating mast cell activation. *Eur. J. Pharmacol.* **300**, 227–236 (1996).
- Facci, L. *et al.* Mast cells express a peripheral cannabinoid receptor with differential sensitivity to anandamide and palmitoylethanolamide. *Proc. Natl Acad. Sci. USA* **92**, 3376–3380 (1995).
- Showalter, V. M., Compton, D. R., Martin, B. R. & Abood, M. E. Evaluation of binding in a transfected cell line expressing a peripheral cannabinoid receptor (CB2): identification of cannabinoid receptor subtype selective ligands. *J. Pharmacol. Exp. Ther.* **278**, 989–999 (1996).
- Munro, S., Thomas, K. L. & Abu-Shaar, M. Molecular characterization of a peripheral receptor for cannabinoids. *Nature* **365**, 61–65 (1993).

26. Landsman, R. S., Burkey, T. H., Consroe, P., Roeske, W. R. & Yamamura, H. I. SR141716A is an inverse agonist at the human cannabinoid CB1 receptor. *Eur. J. Pharmacol.* **334**, R1–R2 (1997).
27. Stella, N., Schweitzer, P. & Piomelli, D. A second endogenous cannabinoid that modulates long-term potentiation. *Nature* **388**, 773–778 (1997).
28. Giuffrida, A. & Piomelli, D. Isotope dilution GC/MS determination of anandamide and other fatty acylethanolamides in rat blood plasma. *FEBS Lett.* **422**, 373–376 (1998).
29. Jaggar, S. I., Hasnie, F. S., Sellatray, S. & Rice, A. S. C. The antihyperalgesic actions of the cannabinoid anandamide and the putative CB2 receptor agonist palmitoylethanolamide in visceral and somatic inflammatory pain. *Pain* (in the press).

Acknowledgements. We thank P. Cimminiello, F. De Seta and V. Piscicelli for experimental assistance, and M. Beltramo, A. Makriyannis, L. Sorrentino and N. Stella for discussion. Supported by Neurosciences Research Foundation (A.G. and D.P.), which receives major support from Novartis, and by MURST (G.L.R. and A.C.).

Correspondence and requests for materials should be addressed to D.P. (e-mail: piomelli@nsi.edu).

Expression of a potassium current in inner hair cells during development of hearing in mice

Corné J. Kros^{*†}, J. Peter Ruppersberg^{‡§} & Alfons Rüsch^{†‡§}

^{*} Department of Physiology, School of Medical Sciences, University of Bristol, University Walk, Bristol BS8 1TD, UK

[†] School of Biological Sciences, University of Sussex, Falmer, Brighton BN1 9QG, UK

[‡] Department of Physiology, University of Tübingen, Gmelinstraße 5, 72076 Tübingen, Germany

[§] Division of Sensory Biophysics, ENT Hospital, University of Tübingen, Röntgenweg 11, 72076 Tübingen, Germany

[†] These authors contributed equally to this work.

Excitable cells use ion channels to tailor their biophysical properties to the functional demands made upon them¹. During development, these demands may alter considerably, often associated with a change in the cells' complement of ion channels^{2–4}. Here we present evidence for such a change in inner hair cells, the primary sensory receptors in the mammalian cochlea. In mice, responses to sound can first be recorded from the auditory nerve and observed behaviourally from 10–12 days after birth; these responses mature rapidly over the next 4 days^{5–8}. Before this time, mouse inner hair cells have slow voltage responses and fire spontaneous and evoked action potentials. During development of auditory responsiveness a large, fast potassium conductance is expressed, greatly speeding up the membrane time constant and preventing action potentials. This change in potassium channel expression turns the inner hair cell from a regenerative, spiking pacemaker into a high-frequency signal transducer.

Substantial changes in the ear's anatomy and in gross cochlear potentials occur around the onset and subsequent rapid maturation of auditory function^{7,9–11}. Very little is known, by contrast, about how the membrane currents of the inner and outer hair cells (IHCs and OHCs) develop. We have studied, under whole-cell voltage clamp, time- and voltage-dependent conductances of IHCs, the cells responsible for signalling the reception of sound to the brain, as a function of development between postnatal day (P) 0 and 68.

Before P11–P12, the outward membrane currents of IHCs rise slowly and are fairly small (Fig. 1a). By P13, the outward currents of all cells have acquired an additional fast component (Fig. 1b) which becomes larger over the next few days, as shown for a P19 cell in Fig. 1c. The current–voltage curves of Fig. 1d (filled symbols) show the appearance and the increase in magnitude of the fast component of the outward current that occurs between P11 and P19; this fast component is largely responsible for the increase in total current (open symbols).

The outward currents recorded in mouse IHCs from a few days after the onset of hearing closely resemble the membrane currents

described before for mature guinea-pig IHCs, which also have a fast and a slow component equated with two different potassium currents, $I_{K,f}$ (probably carried by BK channels^{1,12,13}) and $I_{K,s}$ (a delayed rectifier), respectively^{14,15}. The fast component of the current in mouse IHCs is reversibly suppressed by tetraethylammonium (TEA; half-blocking concentration (IC_{50}) estimated from the dose–response curve as 0.30 ± 0.03 mM; $n = 5$), to which the slow component is comparatively resistant ($IC_{50} = 31 \pm 8$ mM, $n = 4$; measured between P17 and P20). The fast component is also blocked by charybdotoxin with an IC_{50} of 21 ± 2 nM ($n = 13$) and by the apparently selective BK-channel-blocker iberiotoxin¹⁶ ($IC_{50} < 10$ nM; $n = 3$). Taken together, this evidence identifies the fast component in mouse IHCs with $I_{K,f}$. The slow component of the mouse K^+ currents is like $I_{K,s}$ in cells studied after the onset of hearing, but the slow current observed at earlier times, in neonatal cells, appears different¹⁵: it inactivates (slowly) and is considerably more sensitive to block by TEA (IC_{50} 4.6 ± 0.9 mM, $n = 6$; P4–P11); we therefore call it $I_{K,neo}$. The development over time of the different currents is shown in Fig. 2. The total current increases sevenfold between P12 and P16 (Fig. 2a), coincident with the onset and rapid maturation of hearing. Figure 2b shows that this increase is not simply due to the increase in cell capacitance (or membrane area), which is less than twofold, and has a different time course. The increase in total current is mostly contributed by the appearance of $I_{K,f}$ (Fig. 2c): there is an increase in the slow current as well, but it is more gradual (Fig. 2d). In the maturing cochlea, develop-

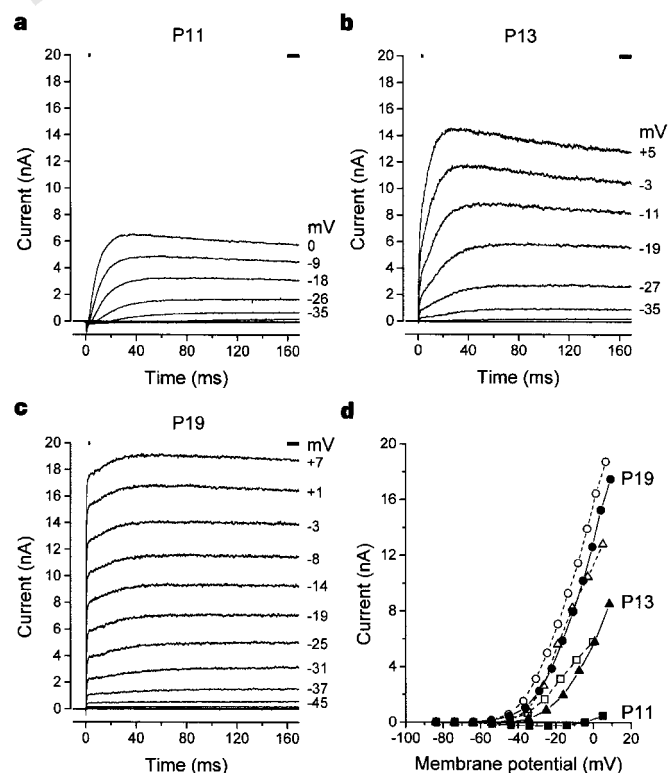


Figure 1 Membrane currents in mouse IHCs of different age. Cells were held at -84 mV, and responses to depolarizations in nominal 10-mV increments were recorded (actual membrane potentials shown by some of the traces). All cells are from the apical coil. Responses are single traces. **a**, At P11, pronounced rapid inward Na^+ currents (over by 2 ms) are followed by outward currents with slow, voltage-dependent kinetics (R_s , 1.0 M Ω , temperature, 24° C). **b**, At P13, outward currents with extremely fast kinetics precede slow outward currents (0.8 M Ω , 22° C). **c**, At P19, the adult pattern of currents is established (2.1 M Ω , 24° C). **d**, Current–voltage curves of the fast component corresponding to $I_{K,f}$ were determined at the time of the short markers in **a–c** (filled symbols, solid lines). Total currents were measured at the long markers (open symbols, dashed lines). Squares: P11; triangles: P13; circles: P19.

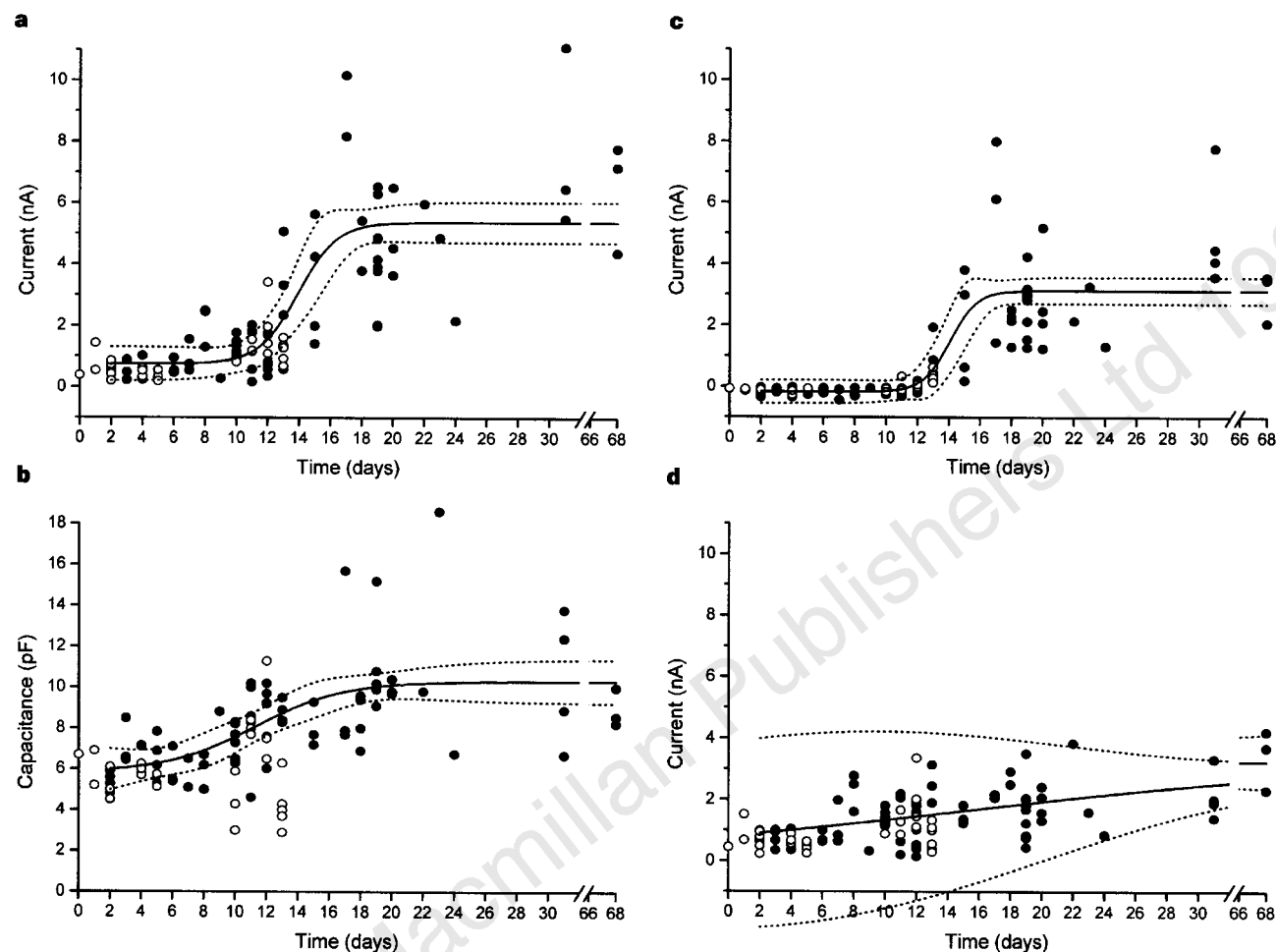


Figure 2 Development of membrane properties of mouse IHCs as a function of postnatal age in a sample of 114 cells. Filled symbols: apical coil IHCs ($n = 85$); open symbols: basal coil IHCs ($n = 29$). **a**, Total, steady-state membrane current measured at -25 mV, elicited from a -84 mV holding potential. **b**, Growth of the membrane capacitance with age. **c**, Fast potassium current ($I_{K,f}$) component. **d**, Slow potassium current component, which is probably a mixture of $I_{K,neo}$ and $I_{K,s}$.

ment is not only a function of time but also of position: various structural maturation processes start at or near the basal end of the cochlea^{10,17}. It is not clear whether the same applies to functional maturation^{10,11}. Our results suggest little or no gradient in the appearance of $I_{K,f}$: the current was first observed in 1 out of 3 IHCs from the basal coil at P11 (but not in 5 apical IHCs), whereas at P12 it was seen in a larger fraction of the cells (4 of 8 cells, apical coil; 1 of 4 cells, basal coil). At P13 all cells expressed $I_{K,f}$ (4 cells, apical coil; 6 cells, basal coil).

The large changes in kinetics and size of the membrane currents that accompany the maturation of auditory function must cause the IHCs to behave very differently *in vivo*. We evaluated the effect of these changes by studying voltage responses of the cells under current clamp at body temperature. Before the onset of hearing, mouse IHCs respond to small current injections with slow and large voltage responses. Depolarizing current injections of more than a few pA result in regenerative, spiking behaviour. The interval between the spikes decreases with increasing current size up to a spike frequency of 15–35 Hz, following which the response turns into a slow oscillation (Fig. 3a). This type of response was seen in nearly all IHCs between P0 and P11. Thirteen out of 31 spiking cells exhibited spontaneous, irregular spikes without current injection at about 2–4 Hz. These responses are due to an interplay of rapidly activating inward Ca^{2+} and Na^{+} currents¹⁸ on the one hand, and the slow delayed rectifier $I_{K,neo}$ on the other, and were seen in hair cells

Fits to the data from the apical coil are according to a sigmoidal logistic growth curve: $C = (C_{max} - C_{min}) / (1 + \exp(-k(t - t_{1/2}))) + C_{min}$, where C is current or capacitance, k is a slope factor, t is time and $t_{1/2}$ is the time where C is halfway between C_{max} and C_{min} . Values for $t_{1/2}$ and k are: **a**, 13.9 d, 0.80 d⁻¹; **b**, 10.3 d, 0.35 d⁻¹; **c**, 14.1 d, 1.16 d⁻¹; **d**, 13.7 d, 0.07 d⁻¹. Dashed lines mark the 95% confidence limits.

from all frequency regions of the cochlea. In various lower vertebrates, with mature auditory systems, such spiking, spontaneous or upon current injection, only occurs in hair cells that can sense very low frequencies^{19–23}. Auditory responses would be modulated and degraded by these electrical characteristics: the slow time constant of 25–70 ms for small depolarizations from the resting potential would cut off periodic responses for frequencies above a few Hz, whereas the spiking prevents intensity coding by means of a graded receptor potential. The expression of the large and fast $I_{K,f}$ improves the frequency response of the IHCs by reducing the membrane time constant to under 1 ms, and at the same time prevents such regenerative behaviour. Figure 3b shows an example of voltage responses at P31, which are comparable to voltage records in mature guinea-pig IHCs¹⁴. There is a small but significant ($P < 0.05$, 2-tailed t -test) difference in the resting potential before (P2–P10: -59.6 ± 1.8 mV; $n = 32$) and after (P19–P68: -65.1 ± 1.5 mV; $n = 19$) the onset of hearing, probably because $I_{K,f}$ and $I_{K,s}$ together provide a larger resting conductance than $I_{K,neo}$.

To ascertain whether the appearance of $I_{K,f}$ is indeed sufficient to abolish spiking, we introduced a simple electronic circuit in the current clamp loop that crudely mimics $I_{K,f}$ (see Methods): 1 nS of artificial $I_{K,f}$ raised the amount of current necessary for spiking to occur from 4 pA to 25 pA for the P2 IHC of Fig. 3a. Introducing a conductance of 2 nS or more abolished spiking (Fig. 3c), despite being only a fraction of the maximum size of $I_{K,f}$: up to 100 nS could

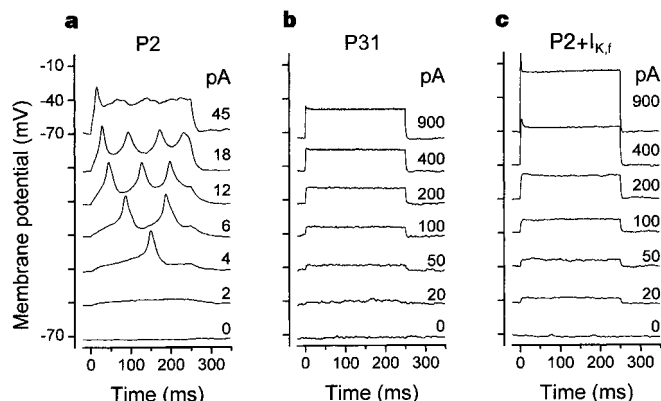


Figure 3 Voltage responses of IHCs to current steps (duration, 250 ms; size next to voltage records). **a**, P2, immature cell. Currents as small as 4 pA elicited slow depolarizations, triggering spikes. **b**, A P31 adult IHC shows rapid, graded voltage responses and has lost the regenerative depolarizations. **c**, Voltage responses of the P2 cell of **a**, with an electronic equivalent of $I_{K,f}$ switched in (4 nS for 0 and 20 pA current steps, 8 nS for other records), resemble those of mature cells. For the largest current steps, progressively more $I_{K,f}$ is activated in the P31 cell, hence the divergence in the voltage responses. For clarity, successive voltage records were shifted by 30 mV, so that unlabelled tick marks represent -70 mV, close to the resting potential. Both cells from apical coil. Temperature was 36°C .

be activated in mature mouse IHCs (measured from tail currents near -50 mV, and about twice as large as previous findings in guinea-pig IHCs¹⁵).

Our findings show that a developmental change in the expression of K^{+} channels contributes to the biophysical maturation of sensory cells in the mammalian inner ear. This adds to the many anatomical, biochemical and physiological changes observed in the peripheral auditory system during the phase of its rapid development, in mice between about P10 and P20 (refs 7, 9). The function of the regenerative spiking of immature IHCs is unknown, but it is conceivable that it drives electrical activity in the central auditory system by means of the afferent synapses already present at birth⁹. In prehatch chicks, cochlear hair cells acquire a calcium-activated potassium current similar to $I_{K,f}$ at a time when rhythmic spontaneous and sound-induced firing in the auditory nerve and brainstem, probably of cochlear origin, disappears^{3,24}. Rhythmic sound-evoked spike patterns have been found at various levels in the central auditory system of kittens¹⁰, and rhythmic spontaneous activity before the onset of hearing has been demonstrated in the auditory nerve and cochlear nucleus of another mammal²⁵. The transducer channels are already functioning in the neonatal mouse well before hearing commences and may act as a noise source to induce spontaneous activity: opening of a single channel would inject 5–10 pA of current²⁶, sufficient to trigger regenerative depolarization. The fast K^{+} conductance first appearing at around P12 shunts the negative slope conductance of voltage-gated Na^{+} and Ca^{2+} channels. The resulting fast time constant of the receptor potential synchronizes Ca^{2+} channel activity with mechanotransduction up to several kHz and enables transmitter release to be phase-locked^{14,27}. The increase in $I_{K,f}$ between P12 and P16 may contribute to the gradual improvement in phase locking observed in auditory nerve fibres during maturation²⁸. □

Methods

IHCs of CD-1 mice were studied either in organotypic cochlear cultures²⁹ (ages P2–P5, where the day of birth is P0), or following acute dissection of the organ of Corti (P0–P68). The mice started to respond behaviourally to a 20 kHz, 90 dB SPL tone burst of 200-ms duration by P12–P14. Membrane currents were studied, at room (20 – 25°C) or near body (35 – 37°C) temperature, by whole-cell patch-clamp, using a List EPC-7 or an Axopatch 200B. The extracellular solution was composed of (mM): 144 NaCl, 0.7 NaH_2PO_4 , 5.8

KCl, 1.3 CaCl_2 , 0.9 MgCl_2 , 5.6 D-glucose, 10 HEPES-NaOH, pH 7.5 at room temperature. Vitamins and amino acids for Eagle's Minimal Essential Medium were added from concentrates (Life Technologies). Patch pipettes were filled with intracellular solution containing (mM): 135 KCl, 0.1 CaCl_2 , 5 EGTA-KOH, 3.5 MgCl_2 , 2.5 Na_2ATP , 5 HEPES-KOH, pH 7.3 at room temperature. Currents under voltage clamp are presented with capacitive transients and linear leak currents subtracted, and all voltages have been corrected for the voltage drop across the uncompensated series resistance, R_s ($2.7 \pm 0.2 \text{ M}\Omega$; $n = 119$), and for liquid junction potentials (-4 or -5 mV) measured between intra- and extracellular solutions. Means are expressed $\pm$ the standard error (s.e.).

Sizes of the potassium currents (Fig. 2) were determined as follows. For experiments at room temperature, the total current was measured around 165 ms from the onset of the depolarizing voltage steps (long markers in Fig. 1a–c) and the fast current at 3 ms (short markers). For body temperature, the time points were 80 ms and 0.7 ms. Current–voltage curves were constructed as in Fig. 1d, and the size of the currents at -25 mV measured; this potential was chosen because all outward conductances were nearly fully activated and the largest currents were still within range of the patch-clamp amplifier (20 nA). The slow current was taken to be the difference between total and fast current. For voltage steps to -25 mV, the activation time constant (single exponential approximation) of $I_{K,f}$ was about 0.6 ms at room temperature and 0.2 ms at body temperature; that of the slow currents $I_{K,s}$ and $I_{K,neo}$ was about 20 ms at room temperature and 4 ms at body temperature. Thus, $I_{K,f}$ was at more than 97% of its steady-state size at the time point for measuring the fast current; the slow potassium currents were at most 15% of steady state: no corrections were attempted for this small contamination.

To obtain realistic voltage responses, all current clamp experiments were conducted at body temperature. An artificial potassium conductance mimicking $I_{K,f}$ could be introduced in the current clamp circuitry of the EPC-7. This was accomplished by comparing the recorded membrane potential of the cell with an electronically set reversal potential of -74 mV, and multiplying the difference with a pre-set conductance of up to 8 nS; the resulting currents were fed back into the current injection input of the amplifier. We did not introduce a time dependence of the artificial conductance, because for the relatively small conductance values used the speed of the IHCs' voltage responses (time constant at least 0.6 ms) is not limited by the extremely rapid onset kinetics of $I_{K,f}$ measured in the mature IHCs (activation time constant at 36°C less than 0.3 ms at all potentials).

Received 19 February; accepted 29 April 1998.

- Hille, B. *Ionic Channels of Excitable Membranes* 2nd edn (Sinauer, Sunderland, MA, 1992).
- O'Dowd, D. L., Ribera, A. B. & Spitzer, N. C. Development of voltage-dependent calcium, sodium and potassium currents in *Xenopus* spinal neurons. *J. Neurosci.* **8**, 792–805 (1988).
- Fuchs, P. A. & Sokolowski, B. H. A. The acquisition during development of Ca-activated potassium currents by cochlear hair cells of the chick. *Proc. R. Soc. Lond. B* **241**, 122–126 (1990).
- Spitzer, N. C. A developmental handshake: neuronal control of ionic currents and their control of neuronal differentiation. *J. Neurobiol.* **22**, 659–673 (1991).
- Mikaelian, D. & Ruben, R. J. Development of hearing in the normal CBA-J mouse. *Acta Otolaryngol.* **59**, 451–461 (1965).
- Shernson, A. & Pujol, R. Age-related changes in the C57BL/6J mouse cochlea. I. Physiological findings. *Dev. Brain Res.* **2**, 65–75 (1982).
- Romand, R. in *Development of Auditory and Vestibular Systems* (ed. Romand, R.) 47–88 (Academic, New York, 1983).
- Ehret, G. in *Development of Auditory and Vestibular Systems* (ed. Romand, R.) 211–237 (Academic, New York, 1983).
- Pujol, R., Lavigne-Rebillard, M. & Lenoir, M. in *Development of the Auditory System* (eds Rubel, E. W., Popper, A. N. & Fay, R. R.) 146–192 (Springer, New York, 1998).
- Walsh, E. J. & Romand, R. in *Development of Auditory and Vestibular Systems 2* (ed. Romand, R.) 161–219 (Elsevier Science, New York, 1992).
- Rübsamen, R. & Lippe, W. R. in *Development of the Auditory System* (eds Rubel, E. W., Popper, A. N. & Fay, R. R.) 193–270 (Springer, New York, 1998).
- Appenrodt, P. & Kros, C. J. Single-channel recordings of large-conductance Ca^{2+} -activated K^{+} channels from guinea-pig inner hair cells. *J. Physiol.* **501**, 147P (1997).
- Kirkwood, N. K. & Kros, C. J. Rapid photorelease of calcium augments the fast potassium conductance of guinea-pig inner hair cells. *J. Physiol.* **504**, 127P (1997).
- Kros, C. J. & Crawford, A. C. Potassium currents in inner hair cells isolated from the guinea-pig cochlea. *J. Physiol.* **421**, 263–291 (1990).
- Kros, C. J. in *The Cochlea* (eds Dallos, P., Popper, A. N. & Fay, R. R.) 318–385 (Springer, New York, 1996).
- Kaczorowski, G. J., Knaus, H.-G., Leonard, R. J., McManus, O. B. & Garcia, M. L. High-conductance calcium-activated potassium channels: structure, pharmacology, and function. *J. Bioenerg. Biomembr.* **28**, 255–267 (1996).
- Mu, M. Y., Chardin, S., Avan, P. & Romand, R. Ontogenesis of rat cochlea. A quantitative study of the organ of Corti. *Dev. Brain Res.* **99**, 29–37 (1997).
- Kros, C. J., Rüsch, A., Richardson, G. P. & Russell, I. J. Sodium and calcium currents in cultured cochlear hair cells of neonatal mice. *J. Physiol.* **473**, 231P (1993).
- Hudspeth, A. J. & Corey, D. P. Sensitivity, polarity, and conductance change in the response of vertebrate hair cells to controlled mechanical stimuli. *Proc. Natl Acad. Sci. USA* **74**, 2407–2411 (1977).

20. Art, J. J. & Fettiplace, R. Variation of membrane properties in hair cells isolated from the turtle cochlea. *J. Physiol.* **385**, 207–242 (1987).
21. Evans, M. G. & Fuchs, P. A. Tetrodotoxin-sensitive, voltage-dependent sodium currents in hair cells from the alligator cochlea. *Biophys. J.* **52**, 649–652 (1987).
22. Fuchs, P. A., Nagai, T. & Evans, M. G. Electrical tuning in hair cells isolated from the chick cochlea. *J. Neurosci.* **8**, 2460–2467 (1988).
23. Sugihara, I. & Furukawa, T. Morphological and functional aspects of two different types of hair cells in the goldfish sacculus. *J. Neurophysiol.* **62**, 1330–1343 (1989).
24. Lippe, W. R. Rhythmic spontaneous activity in the developing avian auditory system. *J. Neurosci.* **14**, 1486–1495 (1994).
25. Gummer, A. W. & Mark, R. F. Patterned neural activity in brain stem auditory areas of a prehearing mammal, the tammar wallaby (*Macropus eugenii*). *NeuroReport* **5**, 685–688 (1994).
26. Géléoc, G. S. G., Lennan, G. W. T., Richardson, G. P. & Kros, C. J. A quantitative comparison of mechanoelectrical transduction in vestibular and auditory hair cells of neonatal mice. *Proc. R. Soc. Lond. B* **264**, 611–621 (1997).
27. Palmer, A. R. & Russell, I. J. Phase-locking in the cochlear nerve of the guinea-pig and its relation to the receptor potential of inner hair cells. *Hear. Res.* **24**, 1–15 (1986).
28. Kettner, R. E., Feng, J.-Z. & Brugge, J. F. Postnatal development of the phase-locked response to low frequency tones of auditory nerve fibres in the cat. *J. Neurosci.* **5**, 275–283 (1985).
29. Russell, I. J. & Richardson, G. P. The morphology and physiology of hair cells in organotypic cultures of the mouse cochlea. *Hear. Res.* **31**, 9–24 (1987).

Acknowledgements. This work was begun at the University of Sussex and was supported by MRC programme grants to I. J. Russell, G. P. Richardson and C.J.K., and a DFG postdoctoral grant to A.R. We thank R. A. Eatock for providing facilities for additional experiments at Baylor College of Medicine, and E. Glowatzki, M. C. Holley, G. P. Richardson and I. J. Russell for commenting on an early version of the manuscript. C.J.K. is a Royal Society University Research Fellow.

Correspondence and requests for materials should be addressed to C.J.K. (e-mail: c.j.kros@bristol.ac.uk).

Mice lacking serum paraoxonase are susceptible to organophosphate toxicity and atherosclerosis

Diana M. Shih^{††}, Lingjie Gu^{††}, Yu-Rong Xia^{††},
Mohamad Navab^{*}, Wan-Fen Li[‡], Susan Hama^{*},
Lawrence W. Castellani^{††}, Clement E. Furlong[§],
Lucio G. Costa[‡], Alan M. Fogelman^{*} & Aldons J. Lusis^{††}

^{*} Division of Cardiology, Department of Medicine, 47-123 CHS, UCLA School of Medicine, Los Angeles, California 90095, USA

[†] Department of Microbiology and Molecular Genetics, Molecular Biology Institute, UCLA, Los Angeles, California 90095, USA

[‡] Department of Environmental Health, University of Washington, Seattle, Washington 98195, USA

[§] Departments of Medicine and Genetics, University of Washington, Seattle, Washington 98195, USA

Serum paraoxonase (PON1) is an esterase that is associated with high-density lipoproteins (HDLs) in the plasma; it is involved in the detoxification of organophosphate insecticides such as parathion and chlorpyrifos^{1–3}. PON1 may also confer protection against coronary artery disease by destroying pro-inflammatory oxidized lipids present in oxidized low-density lipoproteins (LDLs)^{4–8}. To study the role of PON1 *in vivo*, we created *PON1*-knockout mice by gene targeting. Compared with their wild-type littermates, *PON1*-deficient mice were extremely sensitive to the toxic effects of chlorpyrifos oxon, the activated form of chlorpyrifos, and were more sensitive to chlorpyrifos itself. HDLs isolated from *PON1*-deficient mice were unable to prevent LDL oxidation in a co-cultured cell model of the artery wall, and both HDLs and LDLs isolated from *PON1*-knockout mice were more susceptible to oxidation by co-cultured cells than the lipoproteins from wild-type littermates. When fed on a high-fat, high-cholesterol diet, *PON1*-null mice were more susceptible to atherosclerosis than their wild-type littermates.

PON1-knockout mice were produced by targeted disruption of exon 1 of the *PON1* gene (Fig. 1a). Mice homozygous for the mutation were identified by Southern blot analysis (Fig. 1b). The *PON1* mutant allele was transmitted in a mendelian fashion, and *PON1* homozygous mutant mice were normal in their appearance and body weight. We detected no PON1 protein in plasma samples

from homozygous mutant mice by immunoblotting using a polyclonal antiserum prepared against murine PON1 (Fig. 1c). Wild-type (*PON1*^{+/+}) mice on average had 367 units l⁻¹ of paraoxonase activity in plasma, whereas *PON1* homozygous mutant (*PON1*^{-/-}) mice had no detectable paraoxonase activity in their plasma (Fig. 2a). Heterozygous (*PON1*^{+/-}) mice had ~50% of the paraoxonase activity of their wild-type littermates (data not shown). *PON1*-knockout mice also showed an ~85% decrease in plasma arylesterase activity using phenylacetate as substrate, indicating that most, but not all, of this activity is due to PON1 (data not shown).

PON1 is important in the detoxification of organophosphate insecticides. Injection of purified PON1 protects against organophosphate toxicity in rodents^{9–11}, and common polymorphisms in humans of the *PON1* gene dramatically influence the activity of the enzyme towards various organophosphate substrates^{12,13}. To study the role *in vivo* of PON1 in detoxifying these compounds, we examined the sensitivity of *PON1*^{-/-} mice towards chlorpyrifos (CPS) and its oxon (CPO). CPO is a potent cholinesterase inhibitor and is the activated metabolite of CPS; it is a substrate for PON1 hydrolysis. Our *PON1*^{-/-} mice had only 2% of the plasma chlorpyrifos-oxonase activity of *PON1*^{+/+} mice (Fig. 2b), suggesting that PON1 is the predominant enzyme that hydrolyses CPO in the circulation. To investigate its toxicity, three dosages of CPO (1.5, 3, and 6 mg kg⁻¹ body weight) were applied to *PON1* wild-type and null mice. Four hours after treatment, the activity of acetylcholinesterase (AChE) in brain and diaphragm were determined as indicators of toxicity. At the lowest dose (1.5 mg kg⁻¹), CPO did not affect AChE activity in wild-type mice but reduced it by 80 and 74%

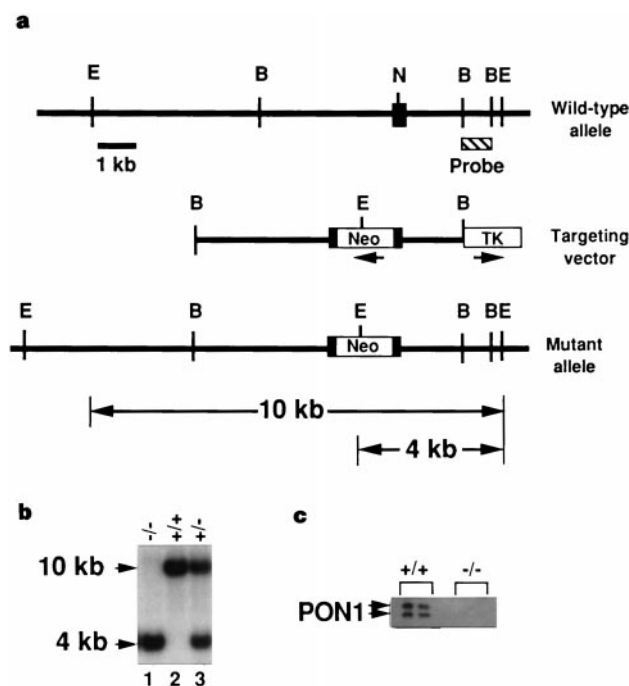


Figure 1 Targeted disruption of the *PON1* gene. **a**, Partial restriction maps of the wild-type *PON1* locus, targeting vector and mutant allele. For gene targeting of the *PON1* locus, exon 1 (filled box) is disrupted by insertion of a positive-selection marker (MC1-Neo). The MC1-TK (thymidine kinase) gene was used for negative selection. The external probe used for Southern blot analysis is shown in the hatched box. B, *Bgl*II; E, *Eco*RI; N, *Nhe*I. **b**, Southern blot analysis of DNA isolated from: lane 1, a mouse homozygous for the targeted allele; lane 2, a mouse homozygous for the wild-type allele; and lane 3, a heterozygous mouse. The wild-type and mutated genes gave 10.0-kb and 4.0-kb fragments, respectively, after *Eco*RI digestion and hybridization with the external probe. **c**, Targeted mice are deficient in PON1 protein. Immunoblotting of PON1 showed the lack of immunoreactive PON1 in the plasma of *PON1*^{-/-} mice.

Table 1 Plasma lipoprotein, lipid and PON1 levels in wild-type, PON1 heterozygous and PON1 null mice

	Total cholesterol	VLDL/LDL cholesterol	HDL cholesterol	Unesterified cholesterol	Triglycerides	Free fatty acids	Relative PON1 activity*
Female, on chow diet							
PON1 ^{+/+} (n = 11)	110 ± 7	13 ± 3†	98 ± 5	17 ± 2	40 ± 7	75 ± 4	100
PON1 ^{+/-} (n = 8)	101 ± 11	11 ± 3‡	90 ± 8	16 ± 3	37 ± 8	77 ± 6	50 ± 4
PON1 ^{-/-} (n = 16)	122 ± 8	19 ± 2†‡	103 ± 6	22 ± 2	66 ± 11	81 ± 6	0 ± 1
Male, on chow diet							
PON1 ^{+/+} (n = 10)	126 ± 8	19 ± 3	107 ± 7	27 ± 3	96 ± 24	79 ± 5	70 ± 9
PON1 ^{+/-} (n = 9)	147 ± 10	21 ± 3	125 ± 9	28 ± 3	93 ± 15	89 ± 5	40 ± 5
PON1 ^{-/-} (n = 16)	136 ± 6	16 ± 2	119 ± 5	27 ± 2	97 ± 14	82 ± 3	0 ± 1
Female, on high-fat diet							
PON1 ^{+/+} (n = 6)	256 ± 22	198 ± 23	58 ± 8	51 ± 4	4 ± 1	36 ± 3	38 ± 4
PON1 ^{-/-} (n = 6)	261 ± 22	197 ± 21	64 ± 10	50 ± 6	3 ± 1	33 ± 1	0 ± 2

Lipid levels are given in mg dl⁻¹ ± s.e. Values for lipid levels on chow diets were from mice of 50% C57BL/6 and 50% 129/SvJ genetic background; lipid levels of mice on an atherogenic, high-fat diet were from animals of 87.5% C57BL/6 and 12.5% 129/SvJ genetic background.

* PON1 activities were determined using arylesterase as substrate and are given as a percentage ± s.e. compared to female wild-type mice fed a chow diet, after subtracting the arylesterase activity in PON1^{-/-} mice. Values are means ± s.e. from sample sizes of between 4 to 8 mice per group. All data were from mice of 87.5% C57BL/6 and 12.5% 129/SvJ genetic background.

† Student's *t*-test: *P* ≤ 0.05.

‡ Student's *t*-test: *P* < 0.05.

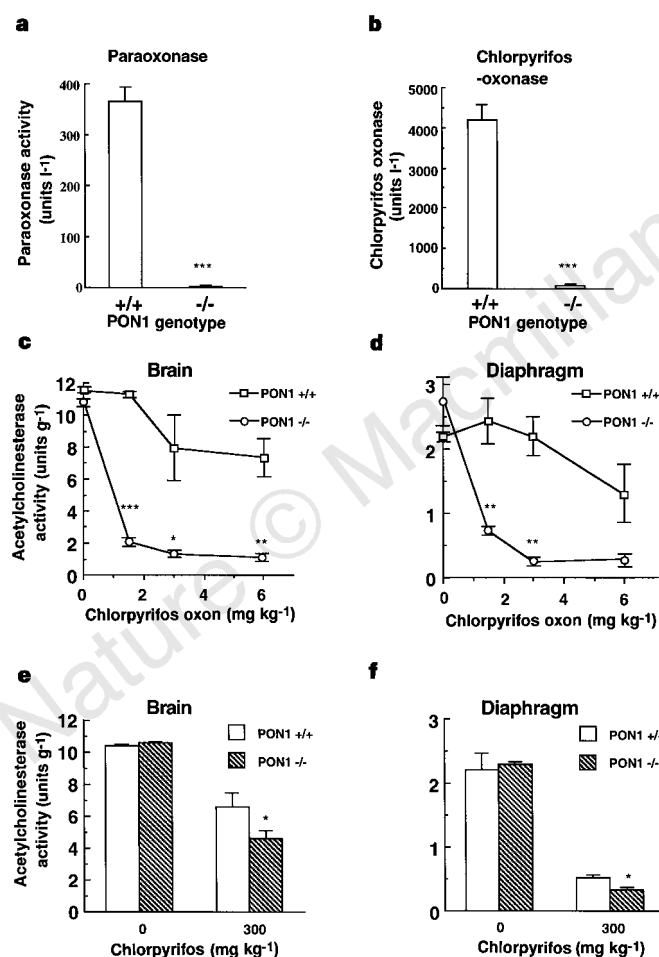


Figure 2 PON1 null mice are more sensitive to the toxic effects of chlorpyrifos oxon and chlorpyrifos. **a**, Plasma paraoxonase, and **b**, chlorpyrifos oxonase activities of wild-type and PON1 null mice. There were 10 and 9 animals in the PON1^{+/+} and PON1^{-/-} groups, respectively. **c**, **d**, Acetylcholinesterase activity in brain (**c**) and diaphragm (**d**) samples from PON1^{+/+} and PON1^{-/-} mice treated with chlorpyrifos oxon. There were 3 animals in each treatment group. **e**, **f**, Acetylcholinesterase activity in brain (**e**) and diaphragm (**f**) samples from PON1^{+/+} and PON1^{-/-} mice treated with chlorpyrifos. There were 4–6 mice in each group. For **a–f**, data are the means obtained from several animals. Error bars indicate standard errors; Student's *t*-test was used for statistical analysis. Asterisks indicate significant differences between PON1^{-/-} and PON1^{+/+} groups at the significance level of **P* < 0.05, ***P* < 0.01, or ****P* < 0.0001, respectively.

in brain and diaphragm, respectively, in PON1-null mice (Fig. 2c, d). At a dose of 3 mg kg⁻¹, CPO inhibited brain AChE activity in PON1^{+/+} animals by 51%, whereas it killed PON1^{-/-} mice within 4 h of exposure. AChE activity in both the brains and diaphragms of PON1^{-/-} mice was significantly lower than that in PON1^{+/+} mice. All of the PON1^{-/-} mice that received 6 mg kg⁻¹ of CPO died within 2 hours of exposure, whereas the PON1^{+/+} mice that received the same dose of CPO survived and their AChE activity was mildly inhibited. PON1^{+/+} mice that received the 6 mg kg⁻¹ or lower doses of CPO showed no symptoms of CPO poisoning. Acute symptoms, including weakness, lacrimation, convulsions and difficulty in breathing, developed in PON1 null mice 1–2 h after CPO exposure.

CPS is not itself a substrate for hydrolysis by PON1 and must first be converted to CPO by cytochrome P450 enzyme systems. We therefore did a toxicology test on PON1 null mice for increased sensitivity to CPS. A high dose of CPS (300 mg kg⁻¹) reduced the AChE activity (in brain and diaphragm tissue) of PON1^{-/-} mice significantly more than in PON1^{+/+} mice (Fig. 2e, f), although the difference in CPS toxicity between wild-type and PON1 null mice was less than the difference in CPO toxicity. We conclude that PON1 null mice are sensitive to the toxic effects of CPO as well as of CPS and that PON1 is responsible for detoxification of organophosphorus pesticides. Therefore, genetic variations in PON1 activity are likely to affect the response of humans to accidental exposure.

As human population studies have revealed an association between PON1 expression and lipoprotein levels⁷, we studied the effects of PON1 deficiency on plasma lipoproteins. We found no significant difference in the amount of lipoproteins in male mice of different PON1 genotypes but, compared with female PON1^{+/+} or PON1^{+/-} mice, female PON1^{-/-} mice had a moderate increase in plasma cholesterol and triglyceride associated with very-low-density lipoproteins (VLDLs) and LDLs (Table 1).

HDLs protect LDLs against oxidation in a co-culture model of the artery wall consisting of a monolayer of aortic endothelial cells overlying the extracellular matrix produced by a layer of smooth muscle cells underneath^{14,15}. In this microenvironment, exogenously added LDLs become modified by oxidation in the subendothelial matrix; the oxidized LDL then induces the synthesis and secretion of monocyte chemoattractant protein-1 (MCP-1) which can be recovered and assayed by its ability to induce monocyte adhesion and transmigration¹⁴. Co-incubation of human LDLs with PON1^{+/+} HDLs decreased monocyte transmigration by 57% compared with incubation with LDL alone (*P* < 0.05) (Fig. 3a). Co-incubation of LDLs with PON1^{+/-} HDLs reduced monocyte transmigration by 38% (data not shown) compared with LDL alone, whereas incubation of LDLs with PON1^{-/-} HDLs did not affect monocyte transmigration, indicating that there is a dose-dependent effect of PON1 expression on the ability of HDL to prevent LDL oxidation. When

$PON1^{-/-}$ HDLs were incubated with co-culture cells without LDL or oxidized phospholipids, they also induced monocyte chemotactic activity, whereas $PON1^{+/+}$ HDLs did not (Fig. 3a). However, the activity induced by $PON1^{-/-}$ HDL alone was significantly lower than in the human LDL plus $PON1^{-/-}$ HDL group ($P < 0.0001$) (Fig. 3a), indicating that HDLs from $PON1$ null mice cannot prevent LDL oxidation and are also pro-inflammatory.

To investigate lipoprotein oxidation and the inflammatory response this induces, we measured the increase in lipid hydroperoxides (LOOH) and in MCP-1 in the medium after treatment of co-cultured cells with HDLs isolated from wild-type and $PON1$ -knockout mice (Fig. 3b–d). We found that pretreatment overnight

of co-cultured cells with $PON1^{+/+}$ HDLs blocked LOOH accumulation in human LDLs (Fig. 3b) and greatly reduced the amount and chemotactic activity of MCP-1 (Fig. 3c, d) in conditioned medium compared with the human LDL-alone group, whereas pretreatment with $PON1^{-/-}$ HDLs did not (Fig. 3b–d).

When $PON1^{-/-}$ HDLs were supplemented with $4 \mu\text{g ml}^{-1}$ purified human $PON1$ protein, paraoxonase activity on these HDLs increased from zero in sham-treated $PON1^{-/-}$ HDL to $9.8 \text{ units ml}^{-1}$. Pretreatment of co-cultured cells with these supplemented $PON1^{-/-}$ HDLs restored the ability of the HDLs to block LOOH accumulation in human LDLs (Fig. 3b) and also decreased the amount of MCP-1 (Fig. 3c) and its chemotactic activity (Fig. 3d) in conditioned medium. Co-cultured cells pretreated with human $PON1$ also prevented LOOH formation in human LDLs (Fig. 3b) and reduced the amount and activity of MCP-1 (Fig. 3c, d) in conditioned medium. These results indicate that $PON1^{-/-}$ HDLs are unable to prevent LDL oxidation because they lack $PON1$.

Platelet-activating factor acetyl hydrolase (PAF-AH), another enzyme carried on HDL, can also destroy biologically active oxidized lipids *in vitro*¹⁶. Although the activity of PAF-AH was not reduced in $PON1$ null mice (data not shown), this HDL was unable to destroy oxidized lipids, indicating that PAF-AH-containing HDLs are ineffective in the absence of $PON1$.

As $PON1$ seems to protect against lipid oxidation, we tested circulating lipoprotein particles for oxidized lipids. The amount of LOOH in each of the HDLs isolated from $PON1^{+/+}$, $PON1^{+/-}$ and $PON1^{-/-}$ mice was 6.2 ± 0.5 , 7.8 ± 0.4 and $8.2 \pm 0.2 \mu\text{g}$ linoleic acid equivalent per mg protein, respectively ($PON1^{+/+}$ versus $PON1^{+/-}$, $P = 0.01$; $PON1^{+/+}$ versus $PON1^{-/-}$, $P = 0.003$). When added by themselves to co-cultured cells, $PON1^{-/-}$ HDLs (without LDLs) stimulated formation of LOOH (Fig. 3b), and MCP-1 levels (Fig. 3c) and chemotactic activity (Fig. 3d) in conditioned medium, whereas $PON1^{+/+}$ HDLs did not (Fig. 3b–d).

Circulating LDLs carry large amounts of antioxidants and previous studies have found no evidence of oxidation. We did not find significant levels of LOOH in freshly isolated LDLs from any $PON1$ genotype either. But in co-culture, LDLs from $PON1$ null mice stimulated more monocyte transmigration than did LDLs from wild-type (3.3-fold increase; $P < 0.0001$; Fig. 3e). Our results suggest that LDLs from $PON1$ null mice are altered in some way that increases their ability to generate oxidized lipids, although the nature of this alteration is unclear.

Our results and previous human genetic epidemiological studies⁸ indicate that reduced $PON1$ may support the development of atherosclerosis. We therefore examined atherosclerotic lesions in the aortae of $PON1$ wild-type, heterozygous and knockout mice fed a high-fat, high-cholesterol diet for 15 weeks. To reduce any variation due to genetic heterogeneity, we backcrossed the $PON1$ null allele for three generations of mice onto the atherosclerosis-susceptible C57BL/6J strain. In mice fed a high-fat diet, the levels of VLDLs/LDLs increased and HDLs and $PON1$ levels were reduced compared with those in mice fed a chow diet (Table 1). The amounts of VLDLs/LDLs and HDLs, as well as of other plasma lipids, were the same in $PON1^{+/+}$ and $PON1^{-/-}$ mice (Table 1), but the $PON1^{-/-}$ mice had significantly larger aortic atherosclerotic lesions (mean lesion area: $13,530 \pm 1,240 \mu\text{m}^2$ per section) than $PON1^{+/+}$ and $PON1^{+/-}$ mice ($7,610 \pm 1,290 \mu\text{m}^2$ per section) ($P < 0.01$) (Fig. 3f). We examined several $PON1^{-/-}$ mice maintained on a chow diet and none showed any evidence of fatty streak lesions, indicating that hyperlipidaemia is required for the development of lesions in these mice (data not shown).

Our results show that $PON1$ null mice are severely compromised with respect to the two processes of inactivating organophosphate poisons and of destroying oxidized lipids. These processes are important in that they determine the response to insecticide poisoning and the susceptibility to atherosclerosis or other inflammatory diseases.

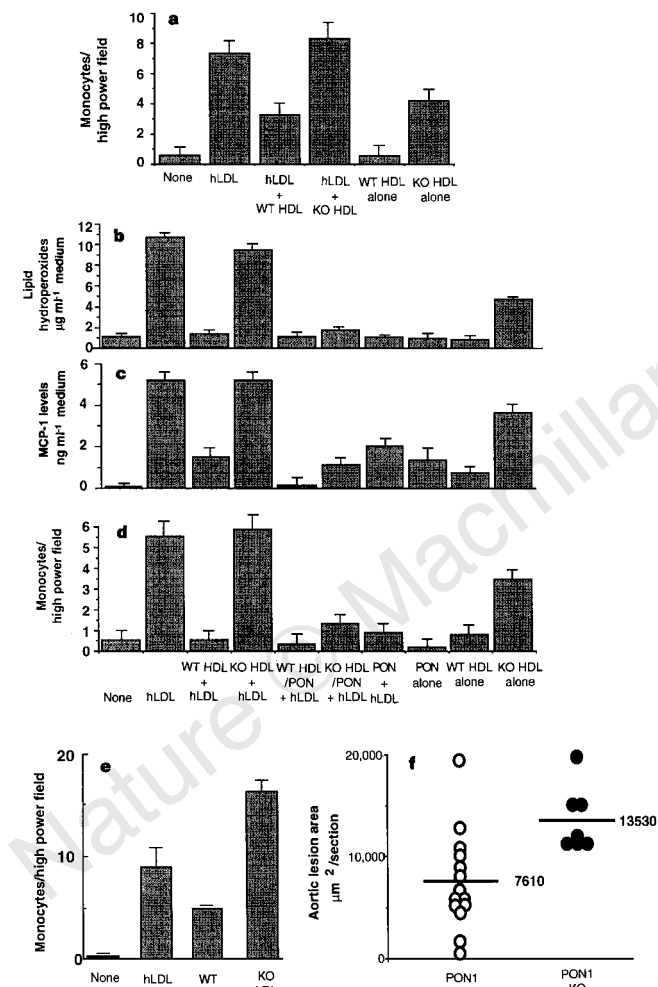


Figure 3 $PON1$ null mice are susceptible to lipoprotein oxidation and atherosclerosis. **a–d**, HDL from $PON1$ null mice fails to protect against LDL oxidation. Co-cultured cells were treated with various combinations of medium alone (none), human LDL (hLDL), wild-type mouse HDL (WT HDL), $PON1$ null mouse HDL (KO HDL), HDL supplemented with purified $PON1$ (HDL/ $PON1$), or purified $PON1$ ($PON1$). In **a**, co-cultures were treated as indicated for 8 h and conditioned medium was tested for monocyte chemotaxis. In **b–d**, co-cultures were pretreated with HDL, with or without supplements of $PON1$, for 18 h and then incubated with fresh medium, with or without hLDL. After incubation, the conditioned media were assayed for LOOH, MCP-1 or monocyte chemotaxis. **e**, Increased susceptibility of LDL from $PON1$ null mice to oxidation. Co-cultured cells were treated with hLDL or with LDL from wild-type (WT LDL) or $PON1$ null (KO LDL) mice, and conditioned media were assayed for monocyte chemotaxis. **f**, Increase in atherosclerosis in $PON1$ null mice. Female $PON1^{+/+}$ or $PON1^{+/-}$ mice ($PON1$ WT/het, open circles) and $PON1^{-/-}$ mice ($PON1$ KO, filled circles) were fed on an atherogenic diet and the lesion area in the proximal aorta was determined. The average lesion sizes are indicated by bars.

Methods

Gene targeting. A gene targeting vector was constructed by first subcloning a 5.3-kb *Bgl*II DNA fragment containing exon 1 of the mouse *PON1* gene and flanking sequences into the *Bam*HI site of the pMCITKpA vector. A neomycin-resistance gene expression cassette, MCINEOpA (Stratagene), was then inserted into the *Nhe*I site of exon 1 of the *PON1* gene fragment. The targeting vector was linearized and electroporated into RW-4 embryonic stem (ES) cells derived from mouse strain 129/SvJ (Genome Systems). G418/gancyclovir-resistant clones were screened by Southern blot analysis and ES cells carrying the disrupted allele were microinjected into blastocysts of mouse strain C57BL/6J to produce chimaeric mice. Chimaeric mice were crossed with C57BL/6J and then intercrossed to produce mice homozygous for the *PON1* mutation.

Activity and immunoblotting of PON1. Plasma PON1 activity towards phenylacetate, paraoxon or chlorpyrifos oxon was determined as described^{16,10}. For immunoblotting of PON1, plasma samples from *PON1* homozygous mutant mice and wild-type littermates were fractionated by SDS-PAGE and transferred onto a nitrocellulose membrane. The membrane was incubated with rabbit anti-mouse PON1 antiserum (1:1,000 dilution), washed, incubated with secondary antibody, and detected using electrochemiluminescence (Amersham).

Toxicity tests and acetylcholinesterase assays. *PON1*^{+/+} and *PON1*^{-/-} mice were treated with various doses of chlorpyrifos oxon or chlorpyrifos dissolved in 20 µl acetone by dermal exposure¹⁰. Four hours after treatment, mice were killed and their brains and diaphragms collected for assay of acetylcholinesterase activity. For mice that died within 4 h of treatment, tissues were collected immediately after death. AChE activity was assayed as described⁹, using acetylthiocholine (ATC) as substrate and 5,5'-dithio-bis(nitrobenzoic acid) (DTNB) as chromogen. AChE activity was expressed as units per g of wet tissue (where one unit is the number of µmol ATC hydrolysed per min).

Plasma lipoproteins. Plasma lipids were determined by enzymatic colorimetric assay¹⁷. LDL and HDL density fractions were isolated from aliquots of pooled plasma from mice of both sexes or from human plasma by ultracentrifugation as described¹⁵. Lipid hydroperoxides in lipoproteins were measured as described¹⁵.

Co-cultures. Co-culturing of human aortic endothelial cells and smooth muscle cells was done as described¹⁴; LDLs were used at a concentration of 350 µg protein ml⁻¹, HDLs at 500 µg protein ml⁻¹, and purified paraoxonase at 4 µg ml⁻¹. When co-cultures were pretreated with HDL or PON1, LOOH levels were determined after 8 h incubation with fresh medium, with or without human LDL; MCP-1 and monocyte chemotaxis were measured after incubation with fresh medium alone for an additional 18 h. MCP-1 was determined by ELISA (Antigenix America) and chemotaxis was assayed, as described¹⁴. For supplementation experiments, murine HDL at 500 µg ml⁻¹ was incubated with purified human paraoxonase at 4 µg ml⁻¹ at 37°C with gentle shaking for 4 h. Unbound paraoxonase was removed by filtering through a 100K M_r-cutoff filter (Amicon). The activity of paraoxonase associated with HDL was determined before and after supplementation; HDLs were then used for co-culture experiments.

Diets and atheromatous-lesion analysis. Mice were fed on either a 6% fat chow diet or, for the atherosclerosis studies, an atherogenic diet containing 15.75% fat, 1.25% cholesterol and 0.5% sodium cholate (Teklad) for 15 weeks¹⁷. Quantification of aortic lesions has been described¹⁷.

Received 27 March; accepted 11 May 1998.

- Murphy, S. D. in *Toxicology: The Basic Science of Poisons* (eds Doull, J., Klassen, C. & Amdur, M.) 357–408 (Macmillan, New York, 1980).
- Tafuri, J. & Roberts, J. Organophosphate poisoning. *Ann. Emerg. Med.* **16**, 193–202 (1987).
- Smolen, A., Eckerson, H. W., Gan, K. N., Hailat, N. & La Du, B. N. Characteristics of the genetically determined allelic forms of human serum paraoxonase/arylesterase. *Drug Metab. Dispos.* **19**, 107–112 (1991).
- Mackness, M. I., Arrol, S. & Durrington, P. N. Paraioxonase prevents accumulation of lipoperoxides in low-density lipoprotein. *FEBS Lett.* **286**, 152–154 (1991).
- Watson, A. D. et al. Protective effect of HDL associated paraioxonase-inhibition of the biological activity of minimally oxidized low density lipoprotein. *J. Clin. Invest.* **96**, 2882–2891 (1995).
- Shih, D. M. et al. Genetic-dietary regulation of serum paraioxonase expression and its role in atherosclerosis in a mouse model. *J. Clin. Invest.* **97**, 1630–1639 (1996).
- Mackness, M. I., Mackness, B., Durrington, P. N., Connelly, P. W. & Hegele, R. A. Paraioxonase: biochemistry, genetics and relationship to plasma lipoproteins. *Curr. Opin. Lipidol.* **7**, 69–76 (1996).
- Heinecke, J. W. & Lusis, A. J. Paraioxonase-gene polymorphisms associated with coronary heart disease: Support for the oxidative damage hypothesis? *Am. J. Hum. Genet.* **62**, 20–24 (1998).

- Costa, L. G. et al. Serum paraioxonase and its influence on paraioxon and chlorpyrifos-oxon toxicity in rats. *Tox. Appl. Pharmacol.* **103**, 66–76 (1990).
- Li, W.-F., Costa, L. G. & Furlong, C. E. Serum paraioxonase status: a major factor in determining resistance to organophosphates. *J. Tox. Envir. Hlth* **40**, 337–346 (1993).
- Li, W.-F., Costa, L. G. & Furlong, C. E. Paraioxonase protects against chlorpyrifos toxicity in mice. *Tox. Lett.* **76**, 219–226 (1995).
- Humbert, R. et al. The molecular basis of the human serum paraioxonase activity polymorphism. *Nature Genet.* **3**, 73–76 (1993).
- Davies, H. G. et al. The effect of the human serum paraioxonase polymorphism is reversed with diazoxon, soman and sarin. *Nature Genet.* **14**, 334–336 (1996).
- Navab, M. et al. Monocyte transmigration induced by modification of low density lipoprotein in coculture of human aortic wall cells is due to induction of monocyte chemotactic protein 1 synthesis and is abolished by high density lipoprotein. *J. Clin. Invest.* **88**, 2039–2046 (1991).
- Navab, M. et al. Mildly oxidized LDL induces an increased apolipoprotein B/paraoxonase ratio. *J. Clin. Invest.* **99**, 2005–2019 (1997).
- Watson, A. D. et al. Effect of platelet activating factor-acetylhydrolase on the formation and action of minimally oxidized low density lipoprotein. *J. Clin. Invest.* **95**, 774–782 (1995).
- Mehrabian, M. et al. Influence of the apoA-II gene locus on HDL levels and fatty streak development in mice. *Arterioscler. Thromb.* **13**, 1–10 (1993).

Acknowledgements. This work was supported by an NIH grant (to A.J.L., A.M.F. and M.N.), an NIEHS grant (to C.E.F.) and an American Heart Association Grant-in-Aid awarded by Greater Los Angeles Affiliate (to D.M.S.). We thank H. Cheroute and K. Williams of UCLA Transgenic Core Facility for blastocyst injections; W. R. Clark and K. M. Lyons for advice in gene targeting work; S. Charugundla for lipid assays; G. P. Hough for MCP-1 assay; and X.-P. Wang and Y.-S. Shi for aortic lesion analysis.

Correspondence and requests for materials should be addressed to A.J.L. (e-mail: jlusis@medicine.medsch.ucla.edu).

Combined effects of angiostatin and ionizing radiation in antitumour therapy

Helena J. Mauceri*, Nader N. Hanna*, Michael A. Beckett*, David H. Gorski†, Mary-Jane Staba‡, Kerri Anne Stellato*, Kevin Bigelow*, Ruth Heimann*, Stephen Gately§, Mohanraj Dhanaball, Gerald A. Soff§, Vikas P. Sukhatmell, Donald W. Kufe¶ & Ralph R. Weichselbaum*

* Department of Radiation and Cellular Oncology, † Department of Surgery, ‡ and Department of Pediatrics, University of Chicago, Chicago, Illinois 60637, USA

§ Department of Medicine, Division of Hematology/Oncology, Northwestern University Medical School, Chicago, Illinois 60611, USA
|| Renal Division, Beth Israel Deaconess Medical Center, Boston, Massachusetts 02115, USA

¶ Dana Farber Cancer Institute and Harvard Medical School, Boston, Massachusetts 02115, USA

Angiogenesis, the formation of new capillaries from pre-existing vessels, is essential for tumour progression^{1–5}. Angiostatin, a proteolytic fragment of plasminogen⁶ that was first isolated from the serum and urine of tumour-bearing mice⁷, inhibits angiogenesis and thereby growth of primary⁸ and metastatic^{7,9,10} tumours. Radiotherapy is important in the treatment of many human cancers, but is often unsuccessful because of tumour cell radiation resistance^{11,12}. Here we combine radiation with angiostatin to target tumour vasculature that is genetically stable and therefore less likely to develop resistance^{13–15}. The results show an antitumour interaction between ionizing radiation and angiostatin for four distinct tumour types, at doses of radiation that are used in radiotherapy. The combination produced no increase in toxicity towards normal tissue. *In vitro* studies show that radiation and angiostatin have combined cytotoxic effects on endothelial cells, but not tumour cells. *In vivo* studies show that these agents, in combination, target the tumour vasculature. Our results provide support for combining ionizing radiation with angiostatin to improve tumour eradication without increasing deleterious effects.

To assess the effects of human angiostatin on primary tumour growth, we treated mice with murine Lewis lung carcinoma (LLC)

tumours with 25 or 50 mg angiostatin per kg per day. The former dose produced a 38% decrease in mean tumour volume and the latter dose reduced mean tumour volume by 54%, as compared with untreated controls (day 9; $P = 0.026$). We selected a dose of 25 mg human angiostatin per kg per day for subsequent studies to allow optimal evaluation of a potential interaction between angiostatin and ionizing radiation.

We examined the effects of human angiostatin and ionizing radiation in LLC tumours and in three human-tumour xenograft models (D54, SQ-20B and PC3). These tumour cell lines, which differ in radiation sensitivities and growth kinetics, are derived from

tumours in which local failure of the primary tumour results in morbidity and mortality. The tumour volumes at the start of treatment ranged from 386 to 1,104 mm³ and represented a tumour burden of 2–5.5% of body weight. The dose of 25 mg human angiostatin per kg per day produced only modest growth inhibition compared with ionizing radiation alone. In contrast, combined treatment with human angiostatin and ionizing radiation produced significant growth inhibition (determined at the nadir; Fig. 1) compared with either treatment alone (Table 1).

In other experiments, we treated mice with LLC tumours with recombinant murine angiostatin. By day 5, 25 mg per kg per day of

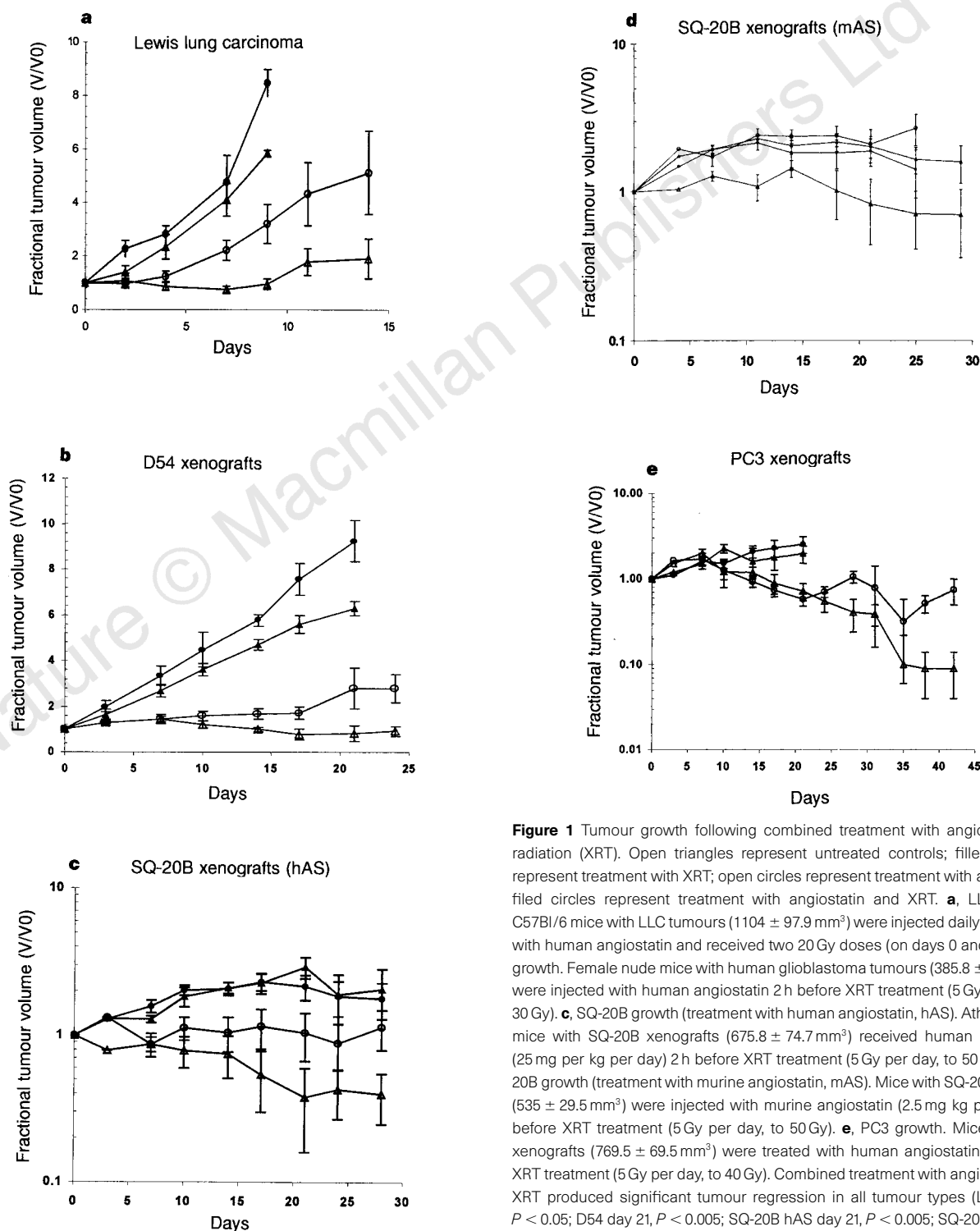


Figure 1 Tumour growth following combined treatment with angiostatin and radiation (XRT). Open triangles represent untreated controls; filled triangles represent treatment with XRT; open circles represent treatment with angiostatin; filled circles represent treatment with angiostatin and XRT. **a**, LLC growth. C57Bl/6 mice with LLC tumours (1104 ± 97.9 mm³) were injected daily for 14 days with human angiostatin and received two 20 Gy doses (on days 0 and 1). **b**, D54 growth. Female nude mice with human glioblastoma tumours (385.8 ± 48.2 mm³) were injected with human angiostatin 2 h before XRT treatment (5 Gy per day to 30 Gy). **c**, SQ-20B growth (treatment with human angiostatin, hAS). Athymic nude mice with SQ-20B xenografts (675.8 ± 74.7 mm³) received human angiostatin (25 mg per kg per day) 2 h before XRT treatment (5 Gy per day, to 50 Gy). **d**, SQ-20B growth (treatment with murine angiostatin, mAS). Mice with SQ-20B tumours (535 ± 29.5 mm³) were injected with murine angiostatin (2.5 mg kg per day) 2 h before XRT treatment (5 Gy per day, to 50 Gy). **e**, PC3 growth. Mice with PC3 xenografts (769.5 ± 69.5 mm³) were treated with human angiostatin 2 h before XRT treatment (5 Gy per day, to 40 Gy). Combined treatment with angiostatin and XRT produced significant tumour regression in all tumour types (LCC day 9, $P < 0.05$; D54 day 21, $P < 0.005$; SQ-20B hAS day 21, $P < 0.005$; SQ-20B mAS day 21, $P < 0.001$; PC3 day 42, $P < 0.001$).

murine angiostatin significantly decreased ($P = 0.007$) mean tumour volume by 29% ($1,680 \pm 111 \text{ mm}^3$) as compared with control tumours ($2,370 \pm 169 \text{ mm}^3$; data not shown). In mice with SQ-20B xenografts (Fig. 1d) treatment with murine angiostatin (2.5 mg per kg per day) and ionizing radiation significantly

Table 1 Summary of fractional tumour volume as a function of treatment

Tumour designation	Control	Radiation	Human angiostatin	Human angiostatin + radiation	
LLC day 9 ($n = 28$)	8.49 ± 0.51	3.21 ± 0.73 (40 Gy)	5.85 ± 0.12	0.96 ± 0.20	$P < 0.05$
D54 day 21 ($n = 20$)	9.25 ± 0.91	2.78 ± 0.89 (30 Gy)	2.94 ± 0.07	0.38 ± 0.16	$P < 0.005$
SQ-20B day 21 ($n = 31$)	2.17 ± 0.50	1.05 ± 0.31 (50 Gy)	2.94 ± 0.07	0.38 ± 0.16	$P < 0.005$
PC3 day 42 ($n = 38$)	*2.55 ± 0.05	0.75 ± 0.35 (40 Gy)	*1.98 ± 0.55	0.09 ± 0.05	$P < 0.001$

The interactive anti-tumour effects of combined treatment with human angiostatin and ionizing radiation are greater than additive when compared with the expected effects of combined treatment. In LLC tumours the mean fractional tumour volume of 3.21 mm³ (radiation-treated group) represents a 62.2% volumetric reduction. If this percentage is analysed with the 31.1% reduction in tumour size in the angiostatin-treated group, then the expected volumetric reduction in the tumour size in the combined treatment group should be 74%. However, the per cent reduction in the combined treatment group is 88.7%, suggesting greater-than-additive treatment effects. Greater-than-additive treatment effects are also observed for D54 tumours (expected, 90.4%; actual, 95.9%), SQ-20B tumours (expected, 34.4%; actual, 82.5%) and PC3 tumours (expected, 77.2%; actual, 96.5%). Data are mean tumour volume $\pm$ s.e.m. per group. Day, nadir of regression; n , total number of animals. * These mice were killed at day 21 because of tumour burden.

reduced mean tumour volume by 64% (day 21; $P < 0.001$) as compared with angiostatin alone (16% reduction) or ionizing radiation alone (18% reduction). These studies, in which we used murine angiostatin at a tenfold lower dose than of human angiostatin, confirmed the interactive antitumour effects of combined treatment with angiostatin and ionizing radiation.

To determine the effects of treatment on tumour neovascularization, we examined representative tissue sections from LLC, D54 and SQ-20B tumours using anti-CD31 antibody and standard immunohistochemical techniques (Fig. 2). The number of LLC and D54 tumour vessels per high-power field was reduced following exposure to combined treatment with human angiostatin and ionizing radiation compared with exposure to all other treatments. The number of vessels per high-power field was also reduced in SQ-20B tumours exposed to combined treatment with angiostatin and ionizing radiation compared with those exposed to ionizing radiation alone ($P = 0.04$). Significant interactive treatment effects were observed in all tumour types (LLC, day 14, $P = 0.06$; D54, day 24, $P = 0.011$; and SQ-20B, day 28, $P = 0.003$; ANOVA).

To explore potential cytotoxic effects of combining angiostatin with ionizing radiation we studied clonogenic survival of human aortic endothelial cells (HAECs) and human umbilical vein endothelial cells (HUVECs) exposed to ionizing radiation. Clonogenic assays demonstrated 30–40% cell killing of HAECs and HUVECs exposed to human angiostatin (10 and 100 ng ml⁻¹). There was no difference in the amount of apoptosis, as measured by staining of HUVEC with 4,6-diamidino-2-phenylindole (DAPI), at 4, 8, 12 or 24 hours of treatment with angiostatin (6.03–9.92% apoptosis) compared with control HUVEC (6.03–10.57%). Inter-

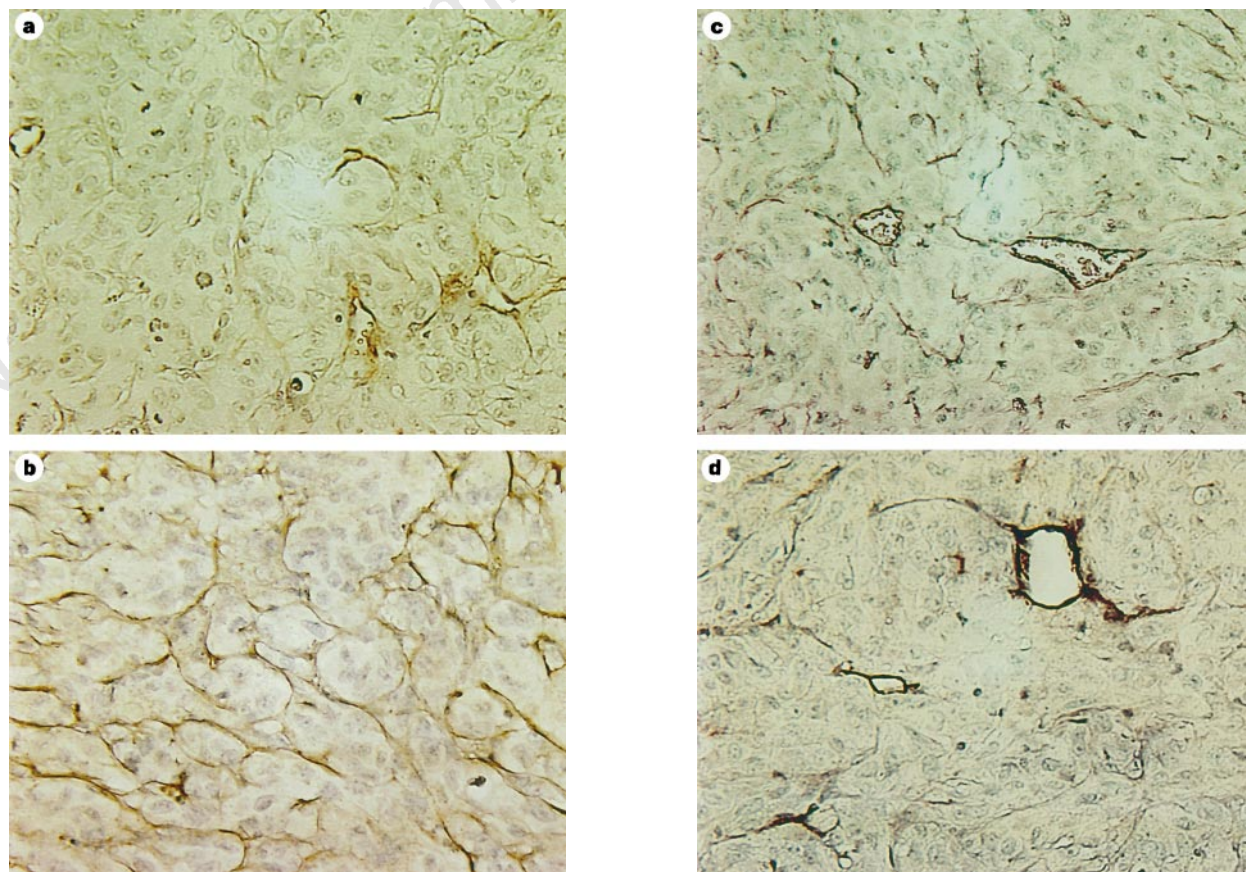


Figure 2 Visualization of the tumour vasculature using anti-CD31 immunohistochemistry. Micrographs are of representative sections from D54 glioblastoma xenografts (day 24) following treatment with human angiostatin and ionizing radiation. Microvessels were visualized in paraffin-embedded tissue sections

using a monoclonal antibody against CD31, and developed using the avidin-biotin peroxidase technique. **a**, Untreated control. **b**, 30 Gy ionizing radiation. **c**, Angiostatin alone. **d**, Angiostatin + 30 Gy ionizing radiation.

active cytotoxicity was observed when HAECs and HUVECs were treated with angiostatin and exposed to 100, 200 or 900 cGy (Fig. 3). Similar results were obtained when we treated bovine aortic endothelial cells (BAECs) with human angiostatin and ionizing radiation. In contrast, no cytotoxicity was observed when tumour cell lines were treated with human angiostatin, and no interactive killing was observed following exposure to angiostatin and ionizing

radiation (Fig. 4). No increase in apoptosis was observed when HUVEC cultures were treated with angiostatin and ionizing radiation (maximum of 21.58% apoptosis at 4 h) compared with ionizing radiation alone (maximum of 22.54% apoptosis at 4 h). These findings indicate that the interactive cytotoxic effects of human angiostatin and ionizing radiation *in vitro* are selective for endothelial cells and not mediated by apoptosis.

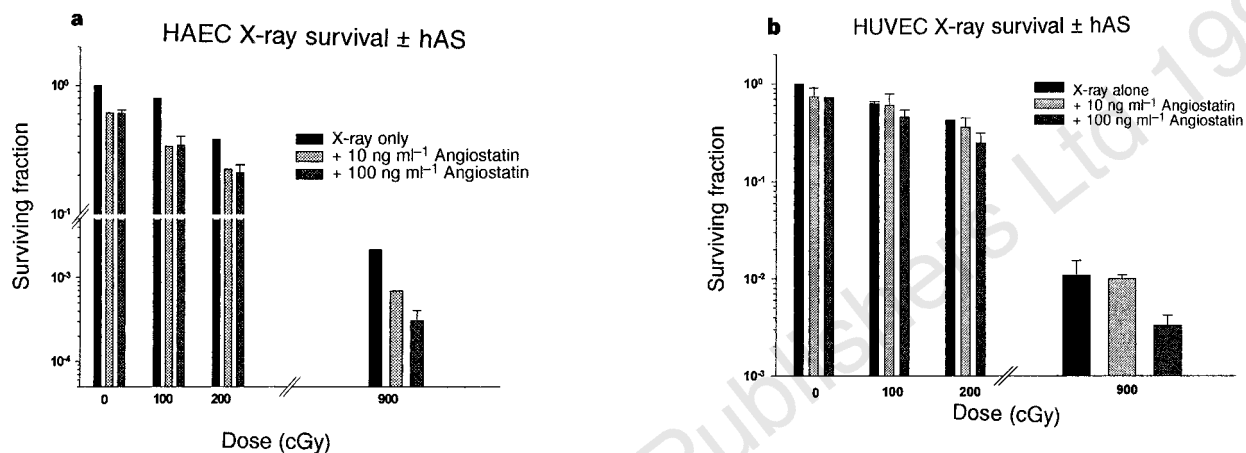


Figure 3 *In vitro* clonogenic survival of **a**, HAECs, and **b**, HUVECs on exposure to X-rays in the presence of human angiostatin (hAS). Cells were plated and exposed to angiostatin 4 hours before a single dose of X-ray irradiation. Colonies

were stained and the surviving fraction was determined. Data represent the mean of two separate experiments $\pm$ s.e.m.

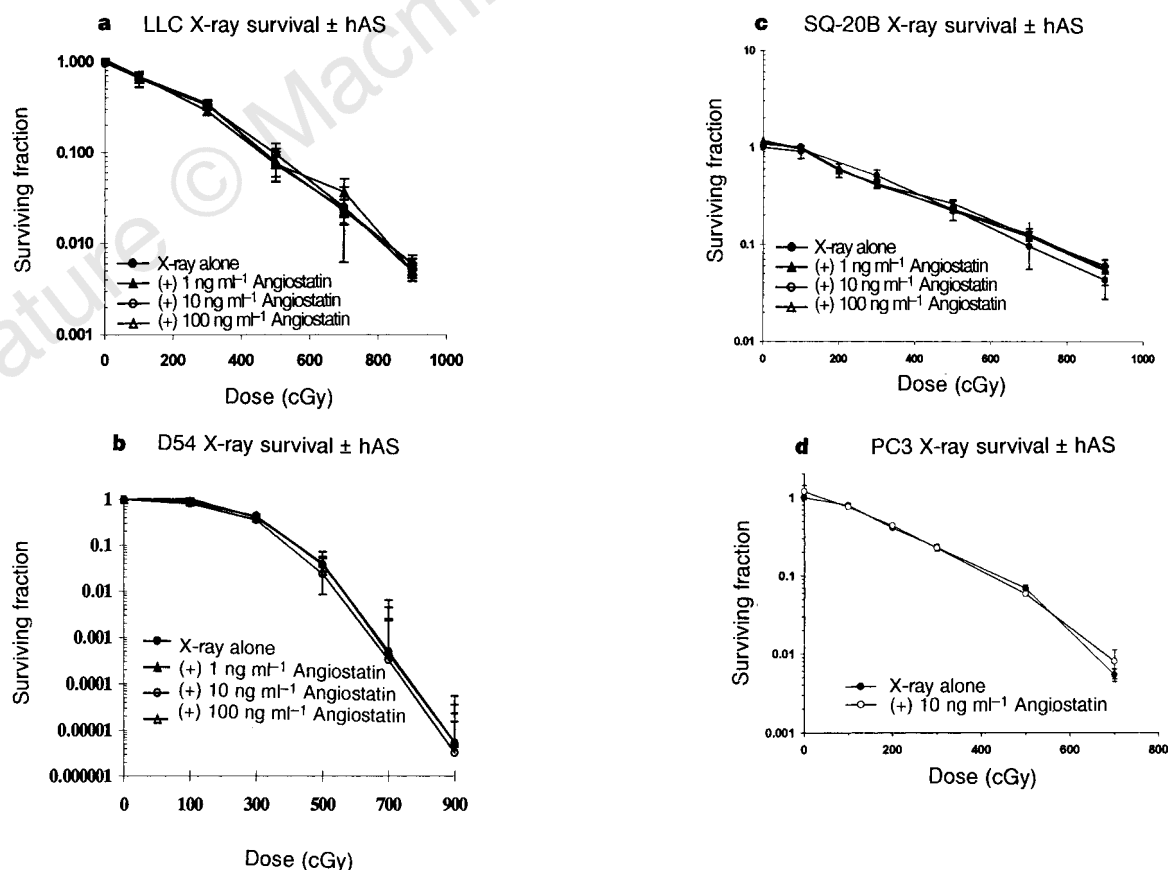


Figure 4 *In vitro* clonogenic survival of **a**, Lewis lung carcinoma cell line (LLC), **b**, human malignant glioma cell line (D54), **c**, human squamous cell carcinoma cell line (SQ-20B), and **d**, human prostate adenocarcinoma cell line (PC3) in response to angiostatin. Cells were plated and exposed to human angiostatin (hAS) 4 h

before a single dose of X-ray irradiation. Colonies were stained and the surviving fraction was determined as described (see Methods). Data are represented as the mean $\pm$ s.e.m.

The interaction of the antitumour effects of angiostatin and ionizing radiation was observed at concentrations of angiostatin that had modest effects on primary tumour growth. We used angiostatin for a relatively short period of time (to coincide with radiotherapy) and used lower doses than those used previously⁸. The relative lack of treatment effects when angiostatin was used as a single agent may be related to the large tumour volumes used here compared with in other studies^{9,10}. Analysis of histological sections of tumours treated with angiostatin and ionizing radiation indicates that newly forming vessels may be the target of the observed antitumour effects. This hypothesis is supported by *in vitro* studies, which show interactive killing of endothelial cell lines by angiostatin and ionizing radiation. Further support for the endothelial cell as a target of the radiation-angiostatin interaction is provided by experiments that show no cytotoxic action or radiosensitizing effects of angiostatin in tumour cell lines. Angiostatin inhibits endothelial cell proliferation *in vitro*; however, this is to our knowledge the first report of clonogenic killing of endothelial cells by angiostatin and synergistic antitumour effects using angiostatin as a radiation modifier. Angiostatin has great potential to enhance the therapeutic ratio in combined-modality cancer treatment. □

Methods

Cell culture. Lewis lung carcinoma cells (LLC-LM) (a gift from J. Folkman) were maintained in DMEM medium (Gibco) with 10% heat-inactivated fetal bovine serum (Intergen) and penicillin/streptomycin (Gibco). Human malignant glioma (D54) cells (a gift from D. D. Bigner) were cultured in 50% DMEM, 50% F-12 medium (Gibco), 7% fetal bovine serum and penicillin/streptomycin. SQ-20B squamous cell carcinoma cells derived from a patient with recurrent squamous cell carcinoma of the larynx were maintained in DMEM: F-12 (3:1), 20% fetal bovine serum, 0.4 µg ml⁻¹ hydrocortisone (Sigma) and penicillin/streptomycin. PC3 prostate adenocarcinoma tumour cells (American Type Culture Collection) were maintained in RPMI-1640 medium (Gibco) containing 10% fetal bovine serum and penicillin/streptomycin. HAECs and HUVECs were maintained in EGM-2 medium (Clonetics Corp.).

Mouse studies. LLC cells were injected subcutaneously (s.c.) into the right hind limb (5 × 10⁵ cells in 100 µl PBS) of C57Bl/6 female mice (Frederick Cancer Research Institute). D54 (1 × 10⁶), SQ-20B (5 × 10⁶) and PC3 cells (2 × 10⁷) were injected into female nude mice. Tumours grew for 17–28 days. Tumour volume was determined by direct measurement with calipers and calculated by the formula (length × width × depth/2). Mice treated with human angiostatin¹⁶ were injected intraperitoneally (i.p.) twice daily (except for SQ-20B) at a total dose of 25 mg per kg per day. Mice with SQ-20B xenografts received a single injection of human angiostatin (25 mg per kg) 2 h before X-ray irradiation. We found no difference in the response to human angiostatin if it was given in one or two injections daily. One group of mice with LLC tumours was injected i.p. twice daily with a dose of 50 mg angiostatin per kg per day. Mice treated with murine angiostatin received a single injection of either 2.5 mg per kg per day (SQ-20B xenografts) or 25 mg per kg per day (LLC tumours). The care and treatment of experimental animals was in accordance with institutional guidelines. Data are reported as percentages of the original (day 0) tumour volume and plotted as fractional tumour volume ± s.e.m.

Angiostatin production. Affinity-purified human angiostatin was generated as a cell-free preparation as described¹⁶. No endotoxin was present in any of the preparations, as shown using the QCL-1000 limulus amoebocyte lysate assay (BioWhittaker).

The gene encoding murine angiostatin was cloned and expressed using a yeast expression system (M.D. and V.P.S., unpublished data). Briefly, a complementary DNA encoding mouse plasminogen (ATCC) was amplified using Vent DNA polymerase. The primers used were designed with the appropriate restriction sites to allow cloning directly into the yeast shuttle plasmid pPICZαA, which contains *EcoRI* and *NotI* restriction sites. Use of this vector permitted secretion of the recombinant protein into the culture medium. Positive clones were grown in 25 ml BMGY medium containing 100 µg ml⁻¹ Zeocin at 30 °C for 18–24 h. The overnight grown culture (A₆₀₀, 2–6) was used to inoculate 2-litre flasks containing 500 ml buffered glycerol

medium. Cells were grown for 2 days at 30 °C (A₆₀₀ 16–20), then centrifuged at 500 r.p.m. for 10 min, and resuspended in 300–400 ml buffered methanol induction medium. The cell-free supernatant was collected on days 2, 3 and 4 and concentrated using ammonium precipitation (70%), dissolved in 50 mM phosphate buffer, pH 7.4, and dialysed at 4 °C. Proteins were purified using a lysine-Sepharose 4B column (Pharmacia) equilibrated with 50 mM phosphate buffer, pH 7.4. Recombinant angiostatin was eluted with 0.2 M L-cysteine-HCl, pH 7.4. Fractions with maximum absorbency were pooled and dialysed for 25 h against PBS, pH 7.4, at 4 °C, with three changes at intervals of 6–8 h. The dialysed sample was further concentrated by ultrafiltration using an Amicon concentrator (YM 10). Protein concentration was determined using the micro BCA assay (BioRad).

Immunohistochemistry. At day 24, mice were killed and tumours were excised and fixed in 10% neutral buffered formalin. After embedding in paraffin, 5 µm sections were cut and tissue sections were mounted. Briefly, sections were dried, deparaffinized and rehydrated. After quenching endogenous peroxidase activity and blocking with bovine serum albumin, slides were incubated at 4 °C overnight with a 1:50 dilution of rat anti-mouse-CD31 monoclonal antibody (Pharmingen), and were then incubated for 1 h with biotinylated rabbit anti-rat immunoglobulin (Vector Laboratories). Localization of blood vessels was visualized using the Vectastain elite ABC kit, Vector PK-6100 (Vector). Slides were dipped in 0.125% osmium tetroxide (Sigma) to enhance positivity and counterstained with 1% methyl green (Trevigen). Ten high-power fields (×400) were examined for each tumour section using a Nikon Microphot-FX microscope equipped with a Sony digital camera. Vessels were counted using Macintosh Image Pro-Plus imaging software.

Clonogenic assay. Endothelial cells (HAECs and HUVECs) were grown in EGM-2 medium (Clonetics). Tumour cell lines (LLC, D54, SQ-20B and PC3) were grown as described. To account for radiation killing, increasing numbers of cells (10² to 5 × 10⁴) were plated in 100-mm tissue-culture dishes. Eighteen hours after plating, angiostatin was added at concentrations of 10 ng ml⁻¹ and 100 ng ml⁻¹. Four hours later, cells were irradiated with doses of 0–900 cGy using a GE Maxitron X-ray generator operating at 250 kV, 26 mA, with a 0.5 mm copper filter, at a dose rate of 118 cGy min⁻¹. Cultures were returned to the incubator for 14–17 days, after which they were stained with crystal violet, colonies were counted, and the surviving fraction was determined. Colonies containing >50 cells were scored as positive.

Statistical analysis. Statistical significance was determined using one-way analysis of variance (ANOVA) or the Kruskal–Wallis test.

Received 1 April; accepted 13 May 1998.

1. Folkman, J. & Shing, Y. Angiogenesis. *J. Biol. Chem.* **267**, 10931–10934 (1992).
2. Fidler, I. J. & Ellis, L. M. The implications of angiogenesis for the biology and therapy of cancer metastasis. *Cell* **79**, 185–188 (1994).
3. Folkman, J. Angiogenesis in cancer, vascular, rheumatoid and other disease. *Nature Med.* **1**, 27–31 (1995).
4. Hanahan, D. & Folkman, J. Patterns and emerging mechanisms of the angiogenic switch during tumorigenesis. *J. Natl Cancer Inst.* **88**, 1091–1092 (1996).
5. Rak, J. W., St Croix, B. D. & Kerbel, R. S. Consequences of angiogenesis for tumor progression, metastasis and cancer therapy. *Anticancer Drugs* **6**, 3–18 (1995).
6. Cao, Y. et al. Krigle domains of human angiostatin. Characterization of the anti-proliferative activity on endothelial cells. *J. Biol. Chem.* **271**, 29461–29467 (1996).
7. O'Reilly, M. Angiostatin: a novel angiogenesis inhibitor that mediates the suppression of metastases by a Lewis lung carcinoma. *Cell* **79**, 315–328 (1994).
8. O'Reilly, M., Holmgren, L., Chen, C. & Folkman, J. Angiostatin induces and sustains dormancy of human primary tumors in mice. *Nature Med.* **2**, 689–692 (1996).
9. Holmgren, L., O'Reilly, M. S. & Folkman, J. Dormancy of micrometastases: balanced proliferation and apoptosis in the presence of angiogenesis suppression. *Nature Med.* **1**, 149–153 (1995).
10. Sim, B. K. et al. A recombinant human angiostatin protein inhibits experimental primary and metastatic cancer. *Cancer Res.* **57**, 1329–1334 (1997).
11. Weichselbaum, R. et al. Radiation-resistant and repair-proficient human tumor cells may be associated with radiotherapy failure in head- and neck-cancer patients. *Proc. Natl Acad. Sci. USA* **83**, 2684–2688 (1986).
12. Weichselbaum, R. R., Beckett, M. A., Dahlberg, W. & Dritschilo, A. Heterogeneity of radiation response of a parent human epidermoid carcinoma cell line and four clones. *Int. J. Radiat. Oncol. Biol. Phys.* **14**, 907–912 (1988).
13. Kerbel, R. S. Inhibition of tumor angiogenesis as a strategy to circumvent acquired resistance to anti-cancer therapeutic agents. *BioEssays* **13**, 31–36 (1991).
14. Kakeji, Y. & Teicher, B. A. Preclinical studies of the combination of angiogenic inhibitors with cytotoxic agents. *Invest. New Drugs* **15**, 39–48 (1997).
15. Teicher, B. A., Sotomayor, E. A. & Huang, Z. D. Antiangiogenic agents potentiate cytotoxic cancer therapies against primary and metastatic disease. *Cancer Res.* **52**, 6702–6704 (1992).
16. Gately, S. et al. The mechanism of cancer-mediated conversion of plasminogen to the angiogenesis inhibitor angiostatin. *Proc. Natl Acad. Sci. USA* **94**, 10868–10872 (1997).

Acknowledgements. We thank S. Hellman and R. Salloum for discussions.

Correspondence and requests for materials and reprints should be addressed to R.R.W. (e-mail: rrw@rover.uchicago.edu).

V(D)J recombinase induction in splenic B lymphocytes is inhibited by antigen-receptor signalling

Marc Hertz^{*†§}, Valérie Kouskoff^{*}, Tetsuya Nakamura[‡] & David Nemazee^{*†}

^{*} National Jewish Medical and Research Center, Division of Basic Sciences, Department of Pediatrics, Denver, Colorado 80206, USA

[†] University of Colorado Health Sciences Center, Department of Immunology, Denver, Colorado 80206, USA

[‡] University of Tokyo, Institute of Medical Science, Department of Infectious Diseases and Applied Immunology, 4-6-1 Shiroganedai, Minato-ku, Tokyo 108, Japan

In lymphocytes, DNA recombinations that generate the antigen-receptor genes can sometimes be reinduced in receptor-bearing cells in a process called receptor editing, which modifies the specificity of the receptor for antigen. In immature B lymphocytes, B-cell antigen receptor (BCR) signalling stimulates immune tolerance by receptor editing^{1–5}. More mature splenic B cells can also be induced to undergo V(D)J recombination, which generates diversity in the immune system, either by immunization with foreign proteins^{6–9} or by stimulation *in vitro* with interleukin-4 and lipopolysaccharide^{8–10}. Here we show that immune tolerance is unlikely to induce V(D)J recombination in mature B cells, because BCR ligation actively inhibits V(D)J recombination induced by interleukin-4 and lipopolysaccharide. Furthermore, immunization of immunoglobulin transgenic mice with ligands of varying avidities for the BCR showed that low-avidity antigen could induce strong V(D)J recombination, whereas non-binding or high-avidity ligands could not. These data suggest that V(D)J recombination induced during the immune response modifies the antigen receptors of B cells with weak, but not strong, reactivity to antigen, potentially rescuing cells with improved receptor affinity and promoting their contribution to the immune response. Thus BCR signalling regulates V(D)J recombination in both tolerance and immunity, but in strikingly different ways.

During B-cell development in the bone marrow, immune tolerance is regulated first by receptor editing and later by the induction of apoptosis¹¹. Despite evidence that apoptosis can eliminate auto-reactive cells in germinal centres^{12–14}, it has not been excluded that the renewed recombinase expression observed in mature splenic B cells^{6–10,15} could represent a recapitulation of this pattern of tolerance-induced receptor editing in self-reactive B cells created during the process of somatic hypermutation. Alternatively, V(D)J recombination may serve as an additional means of somatic diversification in antigen-reactive cells. A more remote possibility is that a subset of immature or pre-B cells present in the spleen is responsible for all of the recombinase activity documented in these previous studies. To distinguish between these possibilities, we induced V(D)J recombination in 3-83 $\mu\delta$ BCR transgenic splenic B cells, and tested the effect of BCR ligation on this induction.

We monitored the frequencies of B cells expressing the immunoglobulin λ light (L) chain because the rearrangement of this locus is normally suppressed in 3-83 $\mu\delta$ transgenic mouse B cells and, unlike endogenously encoded immunoglobulin κ L chains, its protein expression can be readily distinguished from the transgenic immunoglobulin. Treatment of isolated B cells with interleukin-4 (IL-4) and lipopolysaccharide (LPS) stimulated an eightfold

increase in the percentage of λ^+ B cells relative to the control culture (Fig. 1a; compare media and LPS + IL-4 panels) but only a twofold increase in B-cell recovery (Fig. 1b). The absolute number of λ^+ B cells recovered was also greatly increased in cultures containing IL-4 and LPS (Fig. 1c). Importantly, the increase in λ^+ B cells induced by

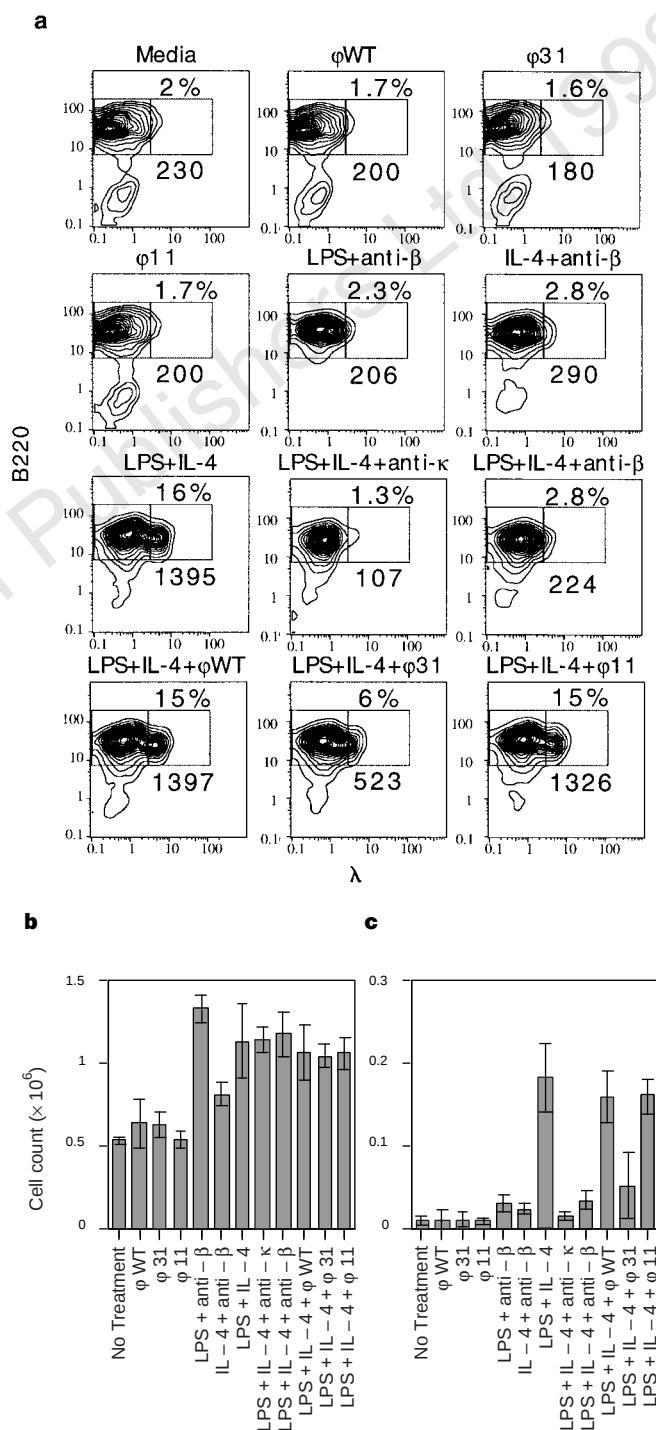


Figure 1 IL-4/LPS treatment of isolated 3-83 $\mu\delta$ transgenic splenic B cells induces an increased percentage and number of λ^+ B cells; this effect can be inhibited by BCR ligation. Data show the effects of various culture conditions on the percentage and absolute number of λ^+ cells after 48 h culture. **a**, Two-colour FACS analysis of treated cultures. The upper number in each histogram shows the percentage, and the lower number the absolute number, of λ^+ /B220⁺ cells per $\sim 10,000$ gated B cells. **b**, Total viable B220⁺ cells. **c**, Total viable λ^+ /B220⁺ cells. Data are mean $\pm$ s.d. of 5 experiments.

§ Present address: M&E Biotech A/S, Kogle Alle 6, DK-2970 Hørsholm, Denmark.

IL-4 and LPS, but not the overall increase in B-cell numbers, was suppressed in cultures supplemented with monoclonal anti- κ -chain antibody (Fig. 1b, c) or with anti-CD79b (Ig β)¹⁶, which should stimulate κ and λ L-chain-bearing cells similarly (Fig. 1b, c). The increase in λ^+ B cells in the cultures containing IL-4 and LPS was also significantly suppressed by inclusion of ϕ 31, which is an M13 phage displaying a peptide specifically bound with high avidity by the 3-83 antibody¹⁷ (Fig. 1a, c). In contrast, BCR non-binding

phage, ϕ WT, and a phage with 100-fold lower avidity for the BCR, ϕ 11, failed to suppress the increase in λ^+ B cells induced by IL-4 and LPS. The phage concentration used was sufficiently high to result in equivalent and plateau BCR occupancy by phages ϕ 31 and ϕ 11 (data not shown). Thus, the ability to suppress the increase in λ^+ cells that was induced by IL-4 and LPS was correlated with avid binding to the BCR (Fig. 1c). Because both anti-BCR monoclonal antibodies and phage ϕ 31 could suppress the response, it is unlikely that the suppression required interaction with Fc receptors on B cells. Treatment with IL-4 and LPS also increased λ^+ B-cell numbers in cultures of cells from non-transgenic mice. Once again, this response could be inhibited with anti-BCR antibodies (data not shown), indicating that these results were not dependent on the use of cells from BCR transgenic mice.

To determine whether preferential proliferation of λ^+ or κ^+ B cells contributed to the observed changes in the percentage of λ^+ cells in cultures containing IL-4 and LPS, we monitored the cell divisions by prelabelling cells with the fluorescent dye carboxyfluorescein diacetate succinimidylester (CFSE) and assessing fluorescence after 48 h of treatment *in vitro*. With each division, CFSE-labelled cells lose approximately half of their fluorescence¹⁸. Although the overall proliferation was increased in all samples treated with LPS alone or in combination with IL-4, the extent of proliferation was similar in cultures containing IL-4 and LPS, regardless of the presence of BCR ligands (Figs 1b and 2a). In addition, BCR ligation did not accelerate the rate of apoptosis in the treated samples (data not shown). Furthermore, the proliferation of λ^+ splenic B cells in these cultures was no greater than that of the λ^- B cells and was not increased by treatment with LPS and IL-4, indicating that the increase induced by LPS and IL-4 in λ^+ B cells was not due to a preferential outgrowth of λ^+ B cells over κ^+ B cells; instead, it suggested that new λ -chain gene rearrangements occurred (Fig. 2b).

We tested the predictions that stimuli inducing λ expression increase *V(D)J* recombinase activity, and that BCR ligation suppresses this response. We found that transgenic B cells treated with LPS and IL-4 had increased levels of the recombinase activator gene *RAG-2* mRNA, as expected from previous work⁸⁻¹⁰, and found that this response was suppressed by BCR ligation (data not shown). Similarly, treatment with LPS and IL-4 induced the formation of λ gene excision product (Fig. 3, compare lanes 1, 7), which was suppressed by BCR ligation (Fig. 3, lanes 8, 9, 11). In contrast, BCR ligation did not disrupt the ability of treatment with LPS and IL-4 to stimulate sterile Ig γ 1 gene transcription (data not shown). Collectively, these data indicate that new immunoglobulin gene rearrangements are induced in splenic B cells treated with LPS and IL-4 and that, unlike in immature bone-marrow B cells, BCR ligation does not stimulate, but rather inhibits, these gene rearrangements.

To test *in vivo* the possible role of BCR signalling in the *V(D)J* recombinant response, we immunized 3-83 μ D transgenic mice with

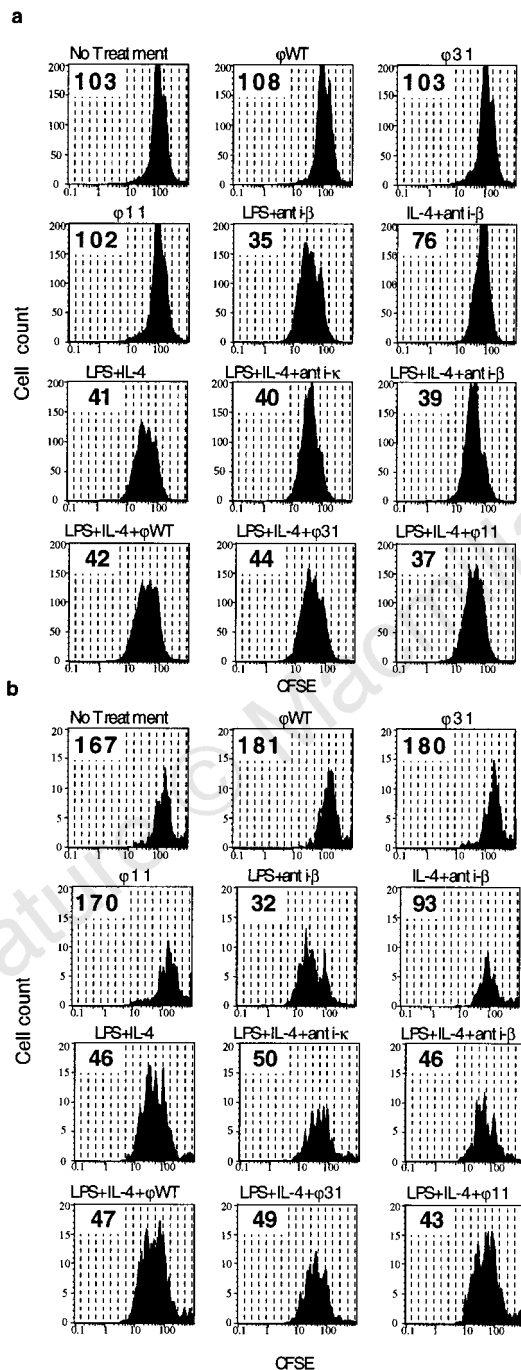


Figure 2 CFSE cell-division analysis of cultured B220⁺ and λ^+ 3-83 μ D transgenic splenic B cells. B cells were stained with CFSE before culture and stimulated for 48 h as indicated in Fig. 1. Broken vertical lines mark successive cell divisions. **a**, FACS analysis of CFSE label in B220⁺ cells at 48 h. **b**, Analysis of CFSE label among gated λ^+ /B220⁺ lymphocytes. The number in the top left corner of each histogram indicates the median fluorescence intensity as an indication of cell division.

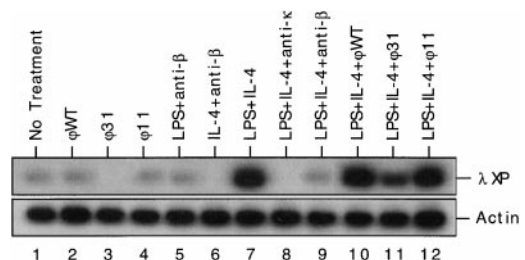


Figure 3 BCR ligation blocks secondary L-chain gene rearrangements induced in purified 3-83 μ D transgenic splenic B cells by IL-4/LPS treatment *in vitro*. Autoradiographs of PCR products revealed by Southern blot analysis of λ L-chain excision products (λ XP, top) and α -actin control gene (Actin, bottom) in cells treated as indicated in Fig. 1.

recombinant phage displaying peptides that bound with varying avidities to the 3-83 BCR¹⁷ (Fig. 4). We found that these mice do not respond to control phage lacking 3-83-reactive peptide, probably because of the low precursor frequency of reactive B cells, but they respond well to phage expressing 3-83-reactive peptide inserts and begin to form germinal centres by 4–5 days (not shown). The frequency of splenic B cells expressing endogenous λ -chains was increased by immunization with the low-avidity phage ϕ 11 (compare Fig. 4a and d), but not by the non-binding phage ϕ WT (Fig. 4b) or the high-avidity phage ϕ 31 (Fig. 4c). This increase in λ^+ B cells in response to ϕ 11 was probably not a result of outgrowth as a consequence of their improved binding to the phage because the 3-83 H-chain paired with a λ -chain has no detectable affinity for the ϕ WT, ϕ 31 or ϕ 11 phage, as measured by enzyme-linked immunosorbent assay (ELISA) (data not shown). Both the frequency and absolute numbers of λ^+ B cells were consistently increased in mice immunized with the low-avidity phage ϕ 11 (Fig. 4e, f). The increase in λ^+ B cells was the result of new $V(D)J$ recombination, as indicated by analysis of the gene-rearrangement DNA excision products in the spleens of immunized mice. λ excision products were rare in the spleens of unimmunized 3-83 $\mu\delta$ transgenic mice (Fig. 4g, lane 4) and were not substantially increased by immunization with control or high-avidity phage (Fig. 4g, lanes 1 and 2, respectively). In contrast, mice immunized with the low-avidity phage had high levels of λ -gene excision products, indicative of newly created λ -gene rearrangements (Fig. 4g, lane 3). It is unlikely that this response is the result of phage stimulating receptor editing directly in the bone marrow of the immunized mice because, if this were the case, phage ϕ 31 should have shown a comparable or better recombinase response. These results are consistent with the idea that B cells with a weak affinity for antigen are able to take up antigen and to initiate

an interaction between helper T cell and B cell, but, because of poor signalling of antigen through their BCRs, cannot suppress the effects of T-cell-derived recombinase-inducing signals, such as CD40L and IL-4. In contrast, B cells with a higher affinity for antigen can be protected from the induction of a recombinase response by stronger BCR-mediated signals.

These data support three important conclusions about $V(D)J$ recombination in peripheral B cells. First, they confirm recent reports documenting the ability of splenic B cells to re-express $V(D)J$ recombinase and to undergo secondary immunoglobulin gene rearrangements^{6–10,15}. Second, the ability of BCR ligation to dampen the expression and activity of the recombinase in splenic B cells indicates that cells undergoing $V(D)J$ recombination express surface immunoglobulin and do not represent pre-B-cell contamination; this was not clear in earlier reports. Finally, the ability of BCR ligation to diminish the recombinase activity in splenic B cells suggests that $V(D)J$ recombination in peripheral B cells is not tolerance-driven, as previously suggested^{6,8–10}. Although RAG gene expression and recombinase activity in germinal-centre centrocytes^{7,9,10} could suggest that self-reactive B cells created during the process of somatic hypermutation may initiate receptor editing, our data and other less direct evidence⁷ make this possibility seem unlikely. The stimuli required to induce the peripheral B-cell recombinase response seem to be T-cell-derived signals, and BCR signalling in this context actively suppresses $V(D)J$ recombination in splenic B cells, whereas BCR signalling in immature surface immunoglobulin-bearing bone-marrow B cells stimulates receptor editing^{3,11}. These striking differences suggest that $V(D)J$ recombination has different roles in the contexts of tolerance and immunity.

During the antibody response, competition among B cells for antigen binding favours the expansion or survival of rare cells that

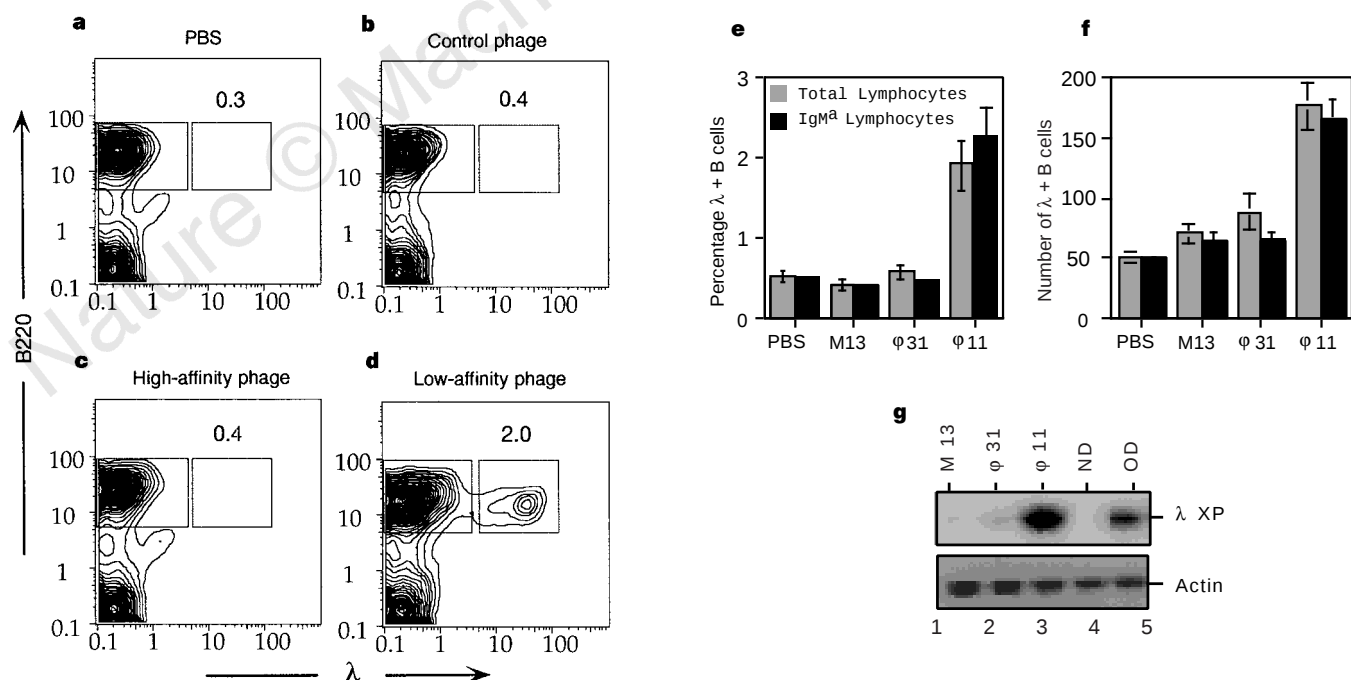


Figure 4 Immunization of 3-83 $\mu\delta$ transgenic mice with phage bearing a low-avidity 3-83 peptide ligand induces secondary immunoglobulin L-chain gene rearrangements *in vivo*. **a–d**, Flow cytometry analysis of splenic lymphocytes 4–5 days after immunization with: **a**, PBS; **b**, control M13 phage, ϕ WT; **c**, phage ϕ 31 expressing high-avidity 3-83 peptide ligand; **d**, phage ϕ 11 expressing low-avidity peptide ligand. Splenic cells (2×10^6) were stained with anti-B220-PE, anti- λ -FITC and anti-IgM^b-biotin. **e**, Percent λ^+ B cells. **f**, Number of λ^+ B cells per 20,000 cells analysed within the lymphocyte gate. Light bars represent total lymphocytes and dark bars represent a subset of IgM^b transgenic H-chain-expressing lymphocytes. In **e** and **f**, data are mean $\pm$ s.d. **g**, Increased endogenous λ gene

rearrangement excision products in splenic B cells from mice immunized with low-avidity peptide ligand. PCR analysis of λ gene rearrangement excision product (XP, top) and α -actin control gene (Actin, bottom) using cell lysate templates made from 3-83 $\mu\delta$ transgenic splenic B cells 4 days after immunization with: control M13 phage (lane 1); phage ϕ 31 high-avidity peptide ligand (lane 2); phage ϕ 11 low-avidity peptide ligand (lane 3); unimmunized 3-83 $\mu\delta$ H-2^d non-deleting (ND) spleen to show untreated background levels of λ excision products (lane 4); and unimmunized 3-83 $\mu\delta$ H-2^b central deleting (CD) spleen as a positive control to show the extent of λ excision products detected in mice undergoing extensive secondary immunoglobulin L-chain gene rearrangements¹ (lane 5).

have a high affinity for antigen¹⁹. Inhibition of receptor revision by the antigen would potentially provide a way of stimulating the expansion of high-affinity clones while allowing lower-affinity cells to improve their affinities. Our phage-immunization results suggest that some BCR signal is required for recruitment to the immune response, but that a strong signal may prevent subsequent receptor revision, at least while antigen quantities last, whereas a weak signal allows receptor revision. It should be remembered, however, that the putative *in vivo* diversification response that we have studied occurs early in, or perhaps before, the germinal-centre response, suggesting that, during the T-cell-dependent response, receptor revision may be initiated before somatic hypermutation.

Within the bone marrow, immature B cells undergo receptor editing to solve a self-tolerance problem, whereas in the periphery receptor revision may contribute to the important processes of intracloonal competition and diversification by stimulating rearrangements in B cells that fail to bind antigen through their BCR. According to this model, receptor revision, like antigen-induced somatic mutation, is likely to create, and potentially select, receptors with affinity to self-tissue. Thus, unlike receptor editing in immature bone-marrow B cells, receptor revision does not solve a self-tolerance problem, but rather seems likely to create one. This situation is reminiscent of T-cell positive selection in the thymus^{20,21}, in which selection for receptors is associated with termination of *V(D)J* recombination, but also with increased risk of autoreactivity, perhaps explaining why both processes are associated with subsequent negative selection by apoptosis^{12–14,22}. □

Methods

Mice and immunizations. We generated 3–83 μ δ transgenic mice²³ by breeding homozygous 3–83 μ δ transgenic males with B10.D2nSn/J (Jackson Laboratories) females. All mice were bred and maintained in the animal care barrier facility at the National Jewish Medical and Research Center. Mice 8–12 weeks old were immunized intraperitoneally with 25 μ g of recombinant M13 phage emulsified in complete Freund's adjuvant (Gibco) and analysed 4–5 days later.

Culture conditions. Splenic B cells were prepared and cultured as described⁸. Briefly, spleens were removed aseptically from mice 8–12 weeks old, cells were dispersed between sterile frosted-glass slides in 10 ml ice-cold sterile HBSS (Gibco) and filtered through sterile cotton to remove debris. Red blood cells were lysed with buffered ammonium chloride. B cells were isolated by anti-Thy-1 and complement depletion of T cells followed by Percoll (Pharmacia) density separation ($\rho = 1.062$) as described²⁴. Purified splenic B cells (10^6 per ml) were cultured in IMDM (Gibco) supplemented with 10% fetal calf serum, 1% sodium pyruvate, 100 U ml⁻¹ penicillin, 100 μ g ml⁻¹ streptomycin, 2 mM L-glutamine and 50 μ M 2- β -mercaptoethanol, and maintained at 37°C with 7% CO₂. Control anti-Ars idotype (5 Ci) (from L. Wysocki), monoclonal anti- κ , 331812 (from T. Rolink), anti-CD79b (Ig- β)¹⁶, HM79 and recombinant M13 phage were added to a final concentration of 10 μ g ml⁻¹, IL-4 (Pharmingen) was added to a final concentration of 500 U ml⁻¹, and LPS (Gibco) was added to a final concentration of 20 μ g ml⁻¹.

Flow cytometric analysis. Cells were analysed by two- or three-colour FACS on a FacsCalibur flow cytometer (Becton Dickinson). Anti-B220-PE (Caltag) was used at a dilution of 1:250; goat anti-mouse λ L-chain-FITC (Caltag) was used at a dilution of 1:1,000; anti-B220-biotin (RA3-3A1, ATCC) was used at a dilution of 1:200; anti-mouse λ -1,2-biotin (Pharmingen) was used at a dilution of 1:200; and anti-IgM-biotin (Pharmingen) was used at a dilution of 1:200. All biotin-conjugated antibodies were revealed with avidin-tricolor (Caltag) at a dilution of 1:50.

Rearrangement polymerase chain reaction (PCR) assay. To generate DNA-containing templates for PCR, 2×10^5 cells were incubated in 200 μ l lysate buffer at 55°C overnight, and amplified by PCR for α -actin or λ L-chain excision products^{1,3}. This assay amplifies and detects the excised DNA episomes generated as the product of *V λ 1-to-J λ 1* rearrangements.

Received 25 March; accepted 5 May 1998.

1. Tiegs, S. L., Russell, D. M. & Nemazee, D. Receptor editing in self-reactive bone marrow B cells. *J. Exp. Med.* **177**, 1009–1020 (1993).
2. Gay, D., Saunders, T., Camper, S. & Weigert, M. Receptor editing: an approach by autoreactive B cells

to escape tolerance. *J. Exp. Med.* **177**, 999–1008 (1993).

3. Hertz, M. & Nemazee, D. BCR ligation induces receptor editing in IgM⁺IgD⁺ bone marrow B cells in vitro. *Immunity* **6**, 429–436 (1997).
4. Chen, C., Prak, E. L. & Weigert, M. Editing disease-associated autoantibodies. *Immunity* **6**, 97–105 (1997).
5. Pelanda, R. *et al.* Receptor editing in a transgenic mouse model: site, efficiency and role in B cell tolerance and antibody diversity. *Immunity* **7**, 765–775 (1997).
6. Han, S., Zheng, B., Schatz, D. G., Spanopoulou, E. & Kelsoe, G. Neoteny in lymphocytes: Rag1 and Rag2 expression in germinal center B cells. *Science* **274**, 2094–2097 (1996).
7. Han, S. *et al.* *V(D)J* recombinase activity in a subset of germinal center B lymphocytes. *Science* **278**, 301–305 (1997).
8. Hikida, M. *et al.* Reexpression of RAG-1 and RAG-2 genes in activated mature mouse B cells. *Science* **274**, 2092–2094 (1996).
9. Hikida, M. & Ohmori, H. Rearrangement of λ light chain genes in mature B cells in vitro and in vivo. Function of reexpressed recombination-activating gene (RAG) products. *J. Exp. Med.* **187**, 795–799 (1998).
10. Papavasiliou, F. *et al.* *V(D)J* recombination in mature B cells: a mechanism for altering antibody responses. *Science* **278**, 298–301 (1997).
11. Melamed, D., Benschop, R. J., Cambier, J. C. & Nemazee, D. Developmental regulation of B lymphocyte immune tolerance compartmentalizes clonal selection from receptor selection. *Cell* **92**, 173–182 (1998).
12. Pulendran, B., Kannourakis, G., Nouri, S., Smith, K. G. & Nossal, G. J. Soluble antigen can cause enhanced apoptosis of germinal-centre B cells. *Nature* **375**, 331–334 (1995).
13. Shokat, K. & Goodnow, C. C. Antigen-induced B-cell death and elimination during germinal-centre immune responses. *Nature* **375**, 334–338 (1995).
14. Han, S., Zheng, B., Dal, P. J. & Kelsoe, G. In situ studies of the primary immune response to (4-hydroxy-3-nitrophenyl)acetyl. IV. Affinity-dependent, antigen-driven B cell apoptosis in germinal centers as a mechanism for maintaining self-tolerance. *J. Exp. Med.* **182**, 1635–1644 (1995).
15. Hikida, M., Mori, M., Kawabata, T., Takai, T. & Ohmori, H. Characterization of B cells expressing recombination activating genes in germinal centers of immunized mouse lymph nodes. *J. Immunol.* **158**, 2509–2512 (1997).
16. Nagata, K. *et al.* The Ig α /Ig β heterodimer on μ -negative proB cells is competent for transducing signals to induce early B cell differentiation. *Immunity* **7**, 559–570 (1997).
17. Kouskoff, V. *et al.* Partial agonist ligands of the B cell receptor induce differential lymphokine production and partial tyrosine phosphorylation (submitted to *J. Exp. Med.*).
18. Lyons, A. B. & Parish, C. R. Determination of lymphocyte division by flow cytometry. *J. Immunol. Methods* **171**, 131–137 (1994).
19. Kelsoe, G. Life and death in germinal centers (redux). *Immunity* **4**, 107–111 (1996).
20. Jameson, S. C., Hogquist, K. A. & Bevan, M. J. Positive selection of thymocytes. *Ann. Rev. Immunol.* **13**, 93–126 (1995).
21. Fink, P. J. & Bevan, M. J. Positive selection of thymocytes. *Adv. Immunol.* **59**, 99–133 (1995).
22. Marrack, P., Hugo, P., McCormack, J. & Kappler, J. Death and T cells. *Immun. Rev.* **133**, 119–129 (1993).
23. Russell, D. M. *et al.* Peripheral deletion of self-reactive B cells. *Nature* **354**, 308–311 (1991).
24. Fleiman, C. M., Hertz, W. M. & Cambier, J. C. Activation of phosphatidylinositol-3' kinase by Src-family kinase SH3 binding to the p85 subunit. *Science* **263**, 1609–1612 (1994).

Acknowledgements. We thank L. Wysocki and T. Rolink for reagents and discussion, and members of D.N.'s lab for critical review of the manuscript and discussion. V. Kouskoff was supported by an EMBO fellowship. This work was supported by grants from the NIH and the Arthritis Foundation.

Correspondence and requests for materials should be addressed to D.N. (e-mail: nemazee@njc.org).

Signals from Ras and Rho GTPases interact to regulate expression of p21^{Waf1/Cip1}

Michael F. Olson, Hugh F. Paterson & Christopher J. Marshall

CRC Centre for Cell and Molecular Biology, Chester Beatty Laboratories, Institute of Cancer Research, 237 Fulham Road, London SW3 6JB, UK

Small GTPases act as molecular switches in intracellular signal-transduction pathways¹. In the case of the Ras family of GTPases, one of their most important roles is as regulators of cell proliferation, and the mitogenic response to a variety of growth factors and oncogenes can be blocked by inhibiting Ras function^{2,3}. But in certain situations, activation of Ras signalling pathways arrests the cell cycle rather than causing cell proliferation^{4–6}. Extracellular signals may trigger different cellular responses by activating Ras-dependent signalling pathways to varying degrees^{7–9}. Other signalling pathways could also influence the consequences of Ras signalling. Here we show that when signalling through the Ras-related GTPase Rho is inhibited, constitutively active Ras induces the cyclin-dependent-kinase inhibitor p21^{Waf1/Cip1} and entry into the DNA-synthesis phase of the cell cycle is blocked. When Rho is active, induction of p21^{Waf1/Cip1} by Ras is suppressed and Ras induces DNA synthesis. Cells that lack p21^{Waf1/Cip1} do not require

Rho signalling for the induction of DNA synthesis by activated Ras, indicating that, once Ras has become activated, the primary requirement for Rho signalling is the suppression of p21^{Waf1/Cip1} induction.

Microinjection of recombinant activated Ras into quiescent NIH 3T3 cells has been shown to induce DNA synthesis¹⁰; however, we found that microinjection of activated Ras (V12 H-Ras) into Swiss 3T3 cells incubated in serum-free medium (serum-starved cells) failed to induce DNA synthesis within 20 h (Fig. 1a) or 44 h (data not shown) after injection. In contrast, if growth-arrested cells were left in medium containing 10% serum (unstarved cells) and then injected with V12 H-Ras, the injected cells began DNA synthesis within 20 h of microinjection (Fig. 1a). Consistent with previous work¹⁰, microinjection of V12 H-Ras into serum-starved NIH 3T3 cells induced DNA synthesis (data not shown).

In some cell types, activation of the Raf/ERK MAP kinase pathway leads to cell-cycle arrest through induction of p21^{Waf1/Cip1} (refs 11, 12). We therefore investigated whether induction of p21^{Waf1/Cip1} could account for the difference in response of Swiss 3T3 cells injected with V12 H-Ras in serum-starved or unstarved conditions. Figure 1b shows that in serum-starved conditions microinjection of

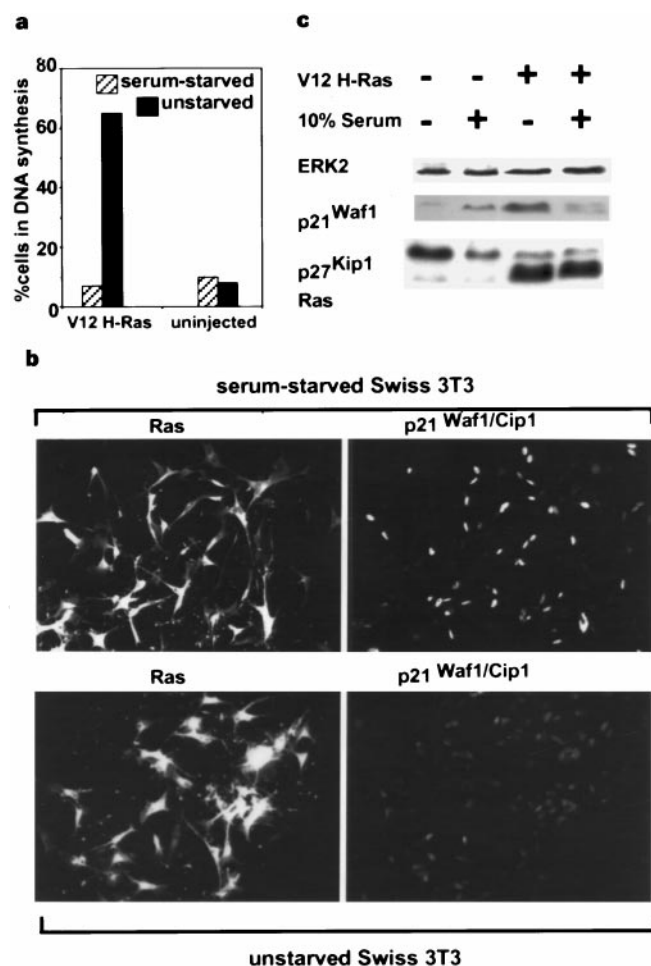


Figure 1 Ras induces p21^{Waf1/Cip1} and fails to induce DNA synthesis in serum-starved Swiss 3T3 cells. **a**, Serum-starved or unstarved Swiss 3T3 cells were microinjected with V12 H-Ras protein and were incubated with 10 μ M BrdU for 16–20 h, fixed and stained for the injection marker and BrdU incorporation¹⁶. **b**, Swiss 3T3 cells were microinjected with V12 H-Ras protein, incubated for 16–20 h, and fixed and stained for the injection marker (left panels) and p21^{Waf1/Cip1} (right panels). **c**, Cell lysates from serum-starved or unstarved Swiss 3T3 cells and V12 H-Ras-expressing Swiss 3T3 cells were immunoblotted for p21^{Waf1/Cip1}, p27^{Kip1}, ERK2 and Ras.

V12 H-Ras induced the expression of p21^{Waf1/Cip1}, whereas in unstarved conditions it did not. In contrast, microinjection of V12 H-Ras into serum-starved NIH 3T3 cells did not induce high levels of p21^{Waf1/Cip1} (Fig. 2b). Microinjection of recombinant activated MAP kinase kinase-1 (MEK-1)¹³ also led to induction of high levels of p21^{Waf1/Cip1} in starved but not unstarved Swiss 3T3 cells (data not shown). Addition of serum to serum-starved cells after injection of V12 H-Ras or activated MEK-1 (data not shown) prevented the induction of p21^{Waf1/Cip1}, showing that an extracellular signal can alter the cellular response to activated Ras or MEK-1.

To determine whether extracellular conditions affect Ras signalling to p21^{Waf1/Cip1} in transformed cells, we prepared lysates from serum-starved and serum-fed Swiss 3T3 control and V12 H-Ras-expressing Swiss cells and analysed the lysates by western blotting.

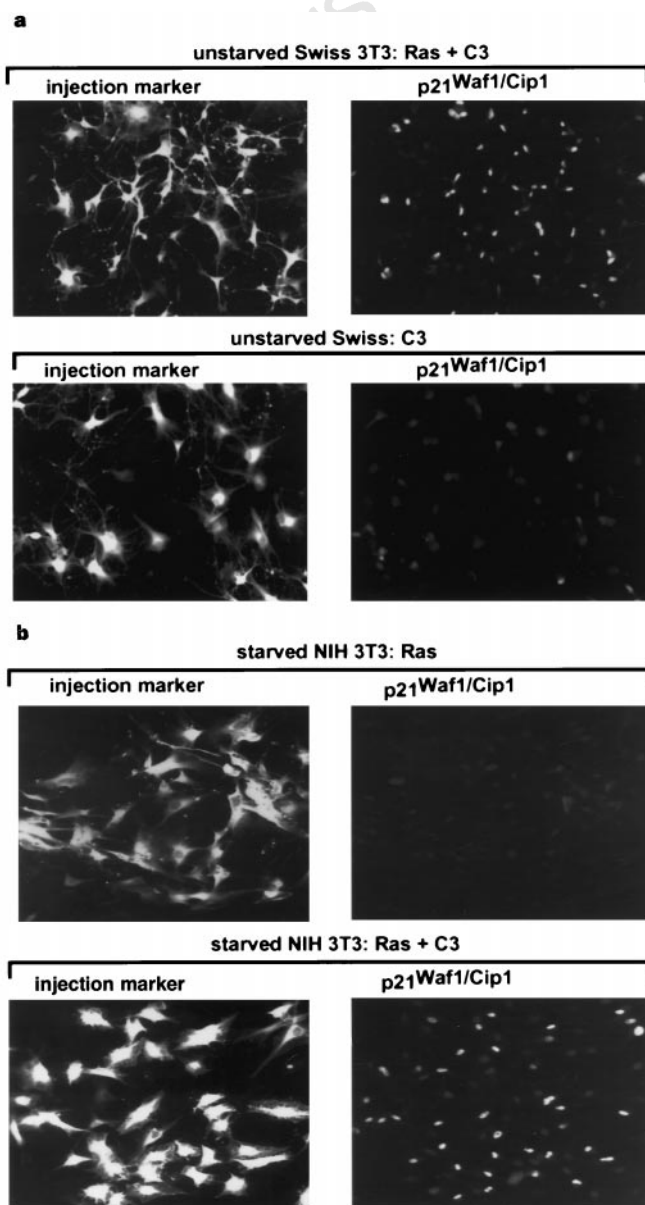


Figure 2 Inhibition of Rho leads to induction of high levels of p21^{Waf1/Cip1} by Ras. **a**, Swiss 3T3 cells in serum-containing medium ('unstarved') were microinjected with V12 H-Ras and C3 exoenzyme or with C3 exoenzyme alone. After 16–20 h, cells were stained for the injection marker (left panels) and p21^{Waf1/Cip1} (right panels). **b**, Serum-starved NIH3T3 cells were injected with V12 H-Ras with and without 10 μ g ml⁻¹ C3 exoenzyme. After 16–20 h, cells were stained as in **a**.

The low levels of p21^{Waf1/Cip1} seen in serum-starved parental Swiss cells were increased in response to serum (Fig. 1c). In contrast, serum-starved V12 H-Ras-expressing cells had elevated p21^{Waf1/Cip1} levels, but these levels were significantly lower when the cells were grown in medium containing serum. Activation of the ERK MAP kinase pathway induced p21^{Waf1/Cip1} in the V12 H-Ras-expressing Swiss cells, as treatment with the MEK inhibitor PD098059 (ref. 12) blocked expression of p21^{Waf1/Cip1} (data not shown). As expected from other studies^{14,15}, levels of another cyclin-dependent-kinase inhibitor, p27^{Kip1}, were reduced in V12 H-Ras-expressing Swiss 3T3 cells compared with parental Swiss cells. Serum did not affect levels of this CDKI in transformed cells (Fig. 1c). These data confirm that Ras influences p21^{Waf1/Cip1} and p27^{Kip1} protein levels in opposite directions: it increases p21^{Waf1/Cip1} levels while decreasing p27^{Kip1}

levels. However, signals from serum can modulate the induction of p21^{Waf1/Cip1} by Ras without affecting the Ras-driven decrease in p27^{Kip1}.

Signalling from the small GTPase Rho is required for serum-induced DNA synthesis¹⁶ and for transformation by Ras^{17–19}. In serum-starved Swiss 3T3 cells, but not in starved NIH 3T3 cells, levels of actin stress fibres are reduced, reflecting reduced Rho activity²⁰. We therefore determined whether serum-induced signalling by Rho prevents Ras from inducing p21^{Waf1/Cip1}. Injection of Ras protein and the specific Rho inhibitor *Clostridium botulinum* C3 exoenzyme²¹ into Swiss 3T3 cells in the presence of serum led to elevated levels of p21^{Waf1/Cip1} (Fig. 2a) compared with those resulting from injection of V12 H-Ras alone (Fig. 1b). In serum-starved NIH 3T3 cells, injection of V12 H-Ras with C3 exoenzyme led to the induction of p21^{Waf1/Cip1} (Fig. 2b). Treatment of serum-starved Swiss 3T3 cells, microinjected with V12 H-Ras, with 0.1 nM *Escherichia coli* cytotoxic necrotizing factor-1 (CNF-1), which selectively activates Rho by deamidating the glutamine at position 63 (refs 22, 23), inhibited p21^{Waf1/Cip1} induction (Fig. 3a). In addition, injection of constitutively active recombinant RhoA protein (L63RhoA; the glutamine at position 63 is replaced with leucine) with V12 H-Ras into Swiss 3T3 cells in serum-starved conditions did not induce p21^{Waf1/Cip1} (data not shown). Thus, Rho signalling antagonizes Ras-induced increases in p21^{Waf1/Cip1} levels.

We then determined whether the suppression of the induction of p21^{Waf1/Cip1} by activated RhoA could allow DNA synthesis (the S phase of the cell cycle) to be induced by activated H-Ras in serum-starved conditions. Cells were microinjected with V12 H-Ras and L63RhoA, or were injected with V12 H-Ras and treated with 0.1 nM CNF-1. Activated RhoA or treatment with CNF-1 allowed most Ras-injected cells to begin new DNA synthesis within 20 h of microinjection (Fig. 3b). Under the conditions used here, neither microinjection of L63RhoA alone nor treatment with CNF-1 alone led to significant DNA synthesis.

To determine whether cells lacking p21^{Waf1/Cip1} still require Rho signalling for S-phase entry induced by activated Ras, we arrested the growth of cultures of mouse embryo fibroblasts (MEFs) from p21^{-/-} homozygous embryos and from wild-type controls and microinjected the cells with V12 H-Ras alone or in conjunction with C3 exoenzyme. As in Swiss 3T3 cells, the level of Rho signalling in serum-starved MEFs was reduced: the cells had fewer actin stress fibres compared with unstarved cells. Microinjection of Ras produced only a 2.5-fold induction of DNA synthesis in wild-type MEFs. Injection of Ras and C3 exoenzyme completely blocked the induction of DNA synthesis by Ras in wild-type cells but had no effect on the sixfold induction of DNA synthesis in p21^{-/-} cells (Fig. 3c). Thus, the principle requirement for Rho signalling in activated-Ras-induced progression into S phase is for the suppression of p21^{Waf1/Cip1} induction.

The increase in p21^{Waf1/Cip1} expression following Raf stimulation of the ERK MAP kinase pathway is mediated, at least in part, by increases in transcription⁸. To determine whether Rho signalling regulates p21^{Waf1/Cip1} levels through effects on transcription, we used NIH 3T3 cell clones containing an integrated p21^{Waf1/Cip1} luciferase reporter²⁴ to reduce the high basal activity of the reporter in transient transfections. As 12-O-tetradecanoylphorbol-13-acetate (TPA) stimulates transcription from the p21^{Waf1/Cip1} promoter^{25,26}, we chose TPA treatment as the stimulus because, unlike other growth factors, TPA does not activate Rho²⁷. After treatment with 40 nM TPA, there was an average of a threefold increase in transcription of the luciferase gene in four independent cell clones (Fig. 4a). However, treatment with 0.1 nM or 1.0 nM CNF-1 to activate Rho markedly inhibited the TPA-induced luciferase activity, indicating that Rho activation antagonized transcription from the p21^{Waf1/Cip1} promoter. These effects of Rho could result from suppression of ERK activation; however, treatment of TPA-stimulated NIH 3T3 cells with 0.1 nM or 1.0 nM CNF-1 did not affect

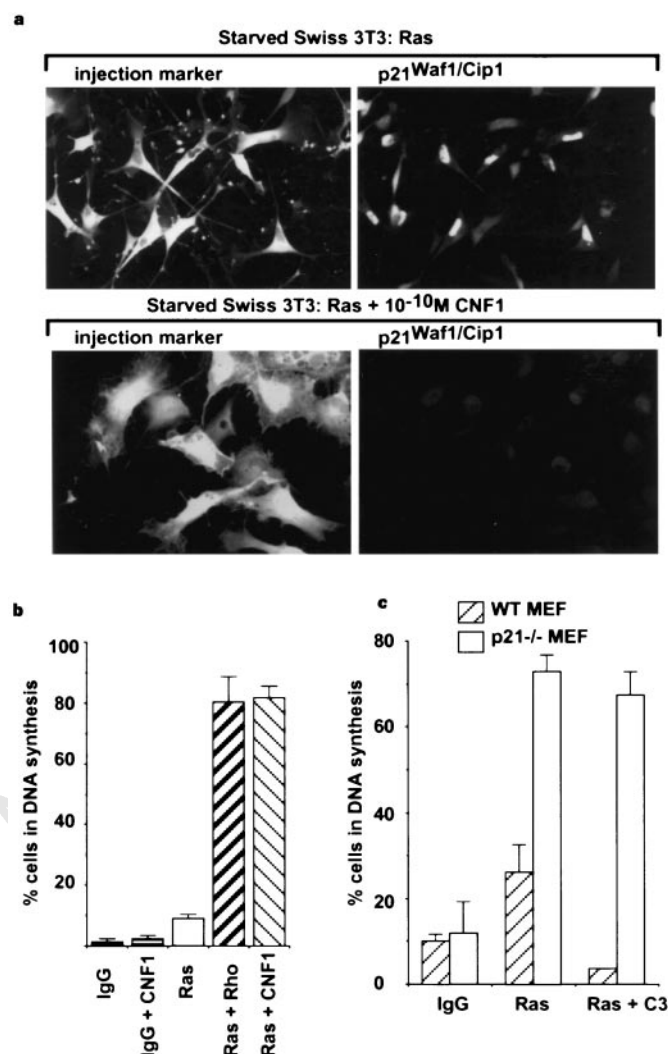


Figure 3 Suppression of p21^{Waf1/Cip1} induction is the major requirement for Rho in Ras-induced DNA synthesis. **a**, Serum-starved Swiss 3T3 cells were microinjected with V12 H-Ras, and, where indicated, 0.1 nM purified CNF-1 (refs 22, 23) was added to the culture medium after injection of Ras. After 16–20 h, cells were stained for the injection marker (left panels) and p21^{Waf1/Cip1} (right panels). **b**, Serum-starved Swiss 3T3 cells were injected with injection marker alone, with V12 H-Ras plus 0.6 mg ml⁻¹ L63RhoA, or with V12 H-Ras. CNF-1 treatment was as in **a**. Cells were incubated for 16–20 h with BrdU and were stained for injection marker and for BrdU incorporation. **c**, Serum-starved p21^{-/-} or wild-type MEFs were microinjected with V12 H-Ras protein or with V12 H-Ras plus C3 exoenzyme, incubated for 16–20 h with BrdU, and stained for injection marker and for BrdU incorporation. **b**, **c**, Data are means $\pm$ s.e.m. from three independent experiments.

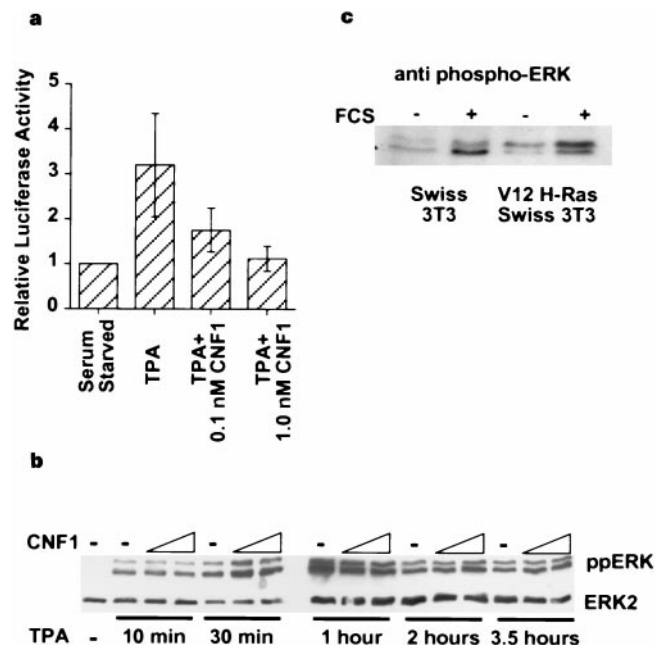


Figure 4 Rho signalling suppresses transcription from the p21^{Waf1/Cip1} promoter. **a**, NIH 3T3 clones stably transfected with the p21^{Waf1/Cip1} promoter reporter plasmid were assayed for luciferase expression following treatment with 40 nM TPA in the presence or absence of 0.1 nM or 1.0 nM CNF-1. **b**, Rho activation does not suppress ERK activation. Serum-starved NIH 3T3 cells were stimulated with 40 nM TPA in the presence or absence of 0.1 nM or 1.0 nM CNF-1 for the times indicated, and were then immunoblotted for activated phosphorylated ERK (ppERK) and total ERK-2. **c**, Serum starvation does not enhance ERK activation in Ras-transformed cells. Serum-starved (-FCS) or unstarved (+FCS) control Swiss 3T3 and V12 H-Ras-expressing Swiss 3T3 cells were immunoblotted for phosphorylated ERK (phospho-ERK).

ERK activation at any time point, as determined by western blotting with an antibody against the dually phosphorylated active form of ERK (Fig. 4b). Stimulation with serum increased ERK activation in parental and V12 H-Ras-transformed Swiss 3T3 cells (Fig. 4c).

The Rho GTPases are essential for Ras-induced oncogenic transformation¹⁷⁻¹⁹ and activated Rho acts synergistically with activated Raf in cell transformation¹⁸. Our results identify a specific target of Rho's actions, which is to counteract the cell-cycle arrest mediated by the induction of p21^{Waf1/Cip1} by Ras. Inhibiting Rho function, which leads to increases in p21^{Waf1/Cip1} levels, may be a new therapeutic strategy against oncogenic Ras-containing tumours. □

Methods

Cell culture and microinjection. Swiss 3T3 cells were grown to confluence in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal calf serum (FCS), and were left in the serum-containing medium (unstarved) or the growth medium replaced with DMEM (serum-starved) for 16–20 h before microinjection. MEFs were growth-arrested by being grown to confluence in DMEM plus 10% FCS, washed three times in DMEM, and incubated in DMEM for 48 h before microinjection. Cells were microinjected on a Zeiss Microinjection Workstation equipped with an Eppendorf 5242 microinjector. Recombinant V12 H-Ras protein was injected at 3 mg ml⁻¹ and C3 exoenzyme was injected at 10 µg ml⁻¹. All microinjections contained rabbit IgG at 1 mg ml⁻¹ as an injection marker. Control microinjections consisted of rabbit IgG alone. For DNA-synthesis measurements, BrdU was added to 10 µM concentration after microinjection, and the cells were incubated for 16–20 h before being fixed and stained with antibody against BrdU¹⁶. DNA synthesis is expressed as the percentage of microinjected cells incorporating BrdU.

Immunofluorescence analysis. Cells were fixed in 4% paraformaldehyde for

15 min and permeabilized in 0.5% Triton X-100. For p21^{Waf1/Cip1} staining, permeabilized cells were blocked with 10% FCS (15 min) and then incubated with undiluted tissue culture supernatant from hybridoma SX-118 (ref. 28) (2 h). The cells were then incubated with a mixture of 1:200 diluted tetramethyl rhodamine isothiocyanate (TRITC)-coupled donkey anti-rabbit IgG (to detect injection marker) and fluorescein isothiocyanate (FITC)-coupled donkey anti-mouse IgG (Jackson ImmunoResearch). For BrdU staining, permeabilized cells were incubated in 1 mg ml⁻¹ DNaseI (Boehringer) for 2 h at 30 °C. They were then treated with rat monoclonal antibody against BrdU and then with TRITC-coupled donkey anti-rabbit IgG and FITC-coupled anti-rat IgG. Confocal images were captured on a BioRad 1024 microscope.

Immunoblotting and kinase assays. 100 µg of cell lysate protein from serum-starved or unstarved Swiss 3T3 cells and V12 H-Ras-expressing Swiss 3T3 cells were run on duplicate 15% SDS-PAGE gels, transferred to nitrocellulose and western-blotted with the p21^{Waf1/Cip1} monoclonal antibody SX-118 (ref. 28), the anti-p27^{Kip1} rabbit polyclonal antibody (Transduction Labs), the anti-ERK2 rabbit polyclonal antibody 122 (ref. 29), the anti-phospho-ERK rabbit polyclonal antibody (Promega), and the anti-pan Ras antibody (Santa Cruz). Protein bands were visualized with horseradish-peroxidase-conjugated anti-mouse or anti-rabbit IgG and electrochemiluminescence (Amersham).

p21^{Waf1/Cip1} reporter cells lines. To generate stable cell lines containing a reporter for transcriptional activation of p21^{Waf1/Cip1}, we transfected a plasmid containing 2.4 kilobases of upstream regulatory sequence of the human p21^{Waf1/Cip1} gene coupled to luciferase²⁴ into NIH 3T3 cells, together with a plasmid encoding neomycin resistance (pSV2-neo). After selection with the neomycin analogue G418, 24 clones were tested for the induction of luciferase expression after phorbol ester treatment (40 nM TPA) to activate the MAP kinase pathway. We selected four clones for further analysis. After 20 h serum starvation, luciferase activity was assayed in untreated or TPA-stimulated cells in the presence or absence of CNF-1, using a Promega luciferase assay kit according to the manufacturer's instructions. Data are shown as the average luciferase activities of the four clones relative to their activities in serum-starved conditions ± s.e.m.

Received 17 April; accepted 21 May 1998.

- Bourne, H. R., Sanders, D. A. & McCormick, F. The GTPase superfamily: a conserved switch for diverse cell functions. *Nature* **348**, 125–132 (1990).
- Mulcahy, L. S., Smith, M. R. & Stacey, D. W. Requirement for ras proto-oncogene function during serum-stimulated growth of NIH 3T3 cells. *Nature* **313**, 241–243 (1985).
- Smith, M. R., DeGubincis, S. J. & Stacey, D. W. Requirement for c-ras proteins during viral oncogene transformation. *Nature* **320**, 540–543 (1986).
- Kohl, N. E. & Ruley, H. E. Role of c-myc in the transformation of REF52 cells by viral and cellular oncogenes. *Oncogene* **2**, 41–48 (1987).
- Ridley, A. J., Paterson, H. F., Noble, M. & Land, H. Ras-mediated cell cycle arrest is altered by nuclear oncogenes to induce Schwann cell transformation. *EMBO J.* **7**, 1635–1645 (1988).
- Serrano, M., Lin, A. W., McCurrach, M. E., Beach, D. & Lowe, S. W. Oncogenic ras provokes premature cell senescence associated with accumulation of p53 and p16INK4A. *Cell* **88**, 593–602 (1997).
- Marshall, C. J. Specificity of receptor tyrosine kinase signalling: transient versus sustained extracellular signal-regulated kinase activation. *Cell* **80**, 179–195 (1995).
- Sewing, A., Wiseman, B., Lloyd, A. C. & Land, H. High-intensity Raf signal causes cell cycle arrest mediated by p21Cip1. *Mol. Cell. Biol.* **17**, 5588–5597 (1997).
- Woods, D. et al. Raf-induced proliferation or cell cycle arrest is determined by the level of Raf activity with arrest mediated by p21Cip1. *Mol. Cell. Biol.* **17**, 5598–5611 (1997).
- Stacey, D. W. & Kung, H.-F. Transformation of NIH 3T3 cells by microinjection of Ha-ras p21 protein. *Nature* **310**, 508–511 (1984).
- Lloyd, A. et al. Cooperating oncogenes converge to regulate cyclin/cdk complexes. *Genes Dev.* **11**, 663–677 (1997).
- Pumiglia, K. M. & Decker, S. J. Cell cycle arrest mediated by the MEK/mitogen-activated protein kinase pathway. *Proc. Natl Acad. Sci. USA* **94**, 448–452 (1997).
- Alessi, D. R. et al. Identification of the sites in MAP kinase kinase-1 phosphorylated by p74raf-1. *EMBO J.* **13**, 1610–1619 (1994).
- Takuwa, N. & Takuwa, Y. Ras activity late in G1 phase required for p27kip1 downregulation, passage through the restriction point, and entry into S phase in growth factor-stimulated NIH 3T3 fibroblasts. *Mol. Cell. Biol.* **17**, 5348–5358 (1997).
- Kawada, M. et al. Induction of p27Kip1 degradation and anchorage independence by Ras through the MAP kinase signalling pathway. *Oncogene* **15**, 629–637 (1997).
- Olson, M. F., Ashworth, A. & Hall, A. An essential role for Rho, Rac, and Cdc42 GTPases in cell cycle progression through G1. *Science* **269**, 1270–1272 (1995).
- Khosravi-Far, R., Solis, P. A., Clark, G. J., Kinch, M. S. & Der, C. J. Activation of Rac1, RhoA and mitogen activated protein kinases is required for Ras transformation. *Mol. Cell. Biol.* **15**, 6443–6453 (1995).
- Qiu, R. G., Chen, J., McCormick, F. & Symons, M. A role for Rho in Ras transformation. *Proc. Natl Acad. Sci. USA* **92**, 11781–11785 (1995).
- Prendergast, G. C. et al. Critical role of Rho in cell transformation by oncogenic Ras. *Oncogene* **10**, 2289–2296 (1995).
- Ridley, A. J. & Hall, A. The small GTP-binding protein rho regulates the assembly of focal adhesions and actin stress fibers in response to growth factors. *Cell* **70**, 389–399 (1992).
- Aktor, K., Mohr, C. & Koch, G. Clostridium botulinum C3 ADP-ribosyltransferase. *Curr. Top. Microbiol. Immunol.* **175**, 115–131 (1992).

22. Flatau, G. *et al.* Toxin-induced activation of the G protein p21 Rho by deamidation of glutamine. *Nature* **387**, 729–733 (1997).
23. Schmidt, G. *et al.* Gln 63 of Rho is deamidated by *Escherichia coli* cytotoxic necrotising factor-1. *Nature* **387**, 725–729 (1997).
24. El-Deiry, W. S. *et al.* WAF1, a potential mediator of p53 tumor suppression. *Cell* **75**, 817–825 (1993).
25. Liu, Y., Martindale, J. L., Gorospe, M. & Holbrook, N. J. Regulation of p21WAF1/CIP1 expression through mitogen-activated protein kinase signaling pathway. *Cancer Res.* **56**, 31–35 (1996).
26. Zeng, Y. X. & el-Deiry, D. W. Regulation of p21WAF1/CIP1 expression by p53-independent pathways. *Oncogene* **12**, 1557–1564 (1996).
27. Ridley, A. J., Paterson, H. E., Johnston, C. L., Diekmann, D. & Hall, A. The small GTP-binding protein rac regulates growth factor-induced membrane ruffling. *Cell* **70**, 401–410 (1992).
28. Fredersdorf, S., Milne, A. W., Hall, P. A. & Lu, X. Characterization of a panel of novel anti-p21Waf1/Cip1 monoclonal antibodies and immunochemical analysis of p21Waf1/Cip1 expression in normal human tissues. *Am. J. Pathol.* **148**, 825–835 (1996).
29. Leever, S. J. & Marshall, C. J. Activation of extracellular signal-regulated kinase, ERK2, by p21 *ras* oncoprotein. *EMBO J.* **11**, 569–574 (1992).

Acknowledgements. We thank X. Lu for the SX-118 anti-p21^{Waf1/Cip1} monoclonal antibody and the p21^{Waf1/Cip1} promoter/luciferase construct; P. Boquet for purified CNF1; S. Hooper for recombinant active MEK1; T. Jacks and G. D'Abaco for mouse embryo fibroblasts; A. Hall for the pGEX-2T L63RhoA plasmid; and L. Feig for the pGEX-2T C3 toxin plasmid. M.F.O. is an Institute of Cancer Research Fellow, and C.J.M. is a Gibb Fellow of the Cancer Research Campaign.

Correspondence and requests for materials should be addressed to C.J.M. (e-mail: chrism@icr.ac.uk).

Structure of dehydroquinase synthase reveals an active site capable of multistep catalysis

Elisabeth P. Carpenter*, Alastair R. Hawkins†, John W. Frost‡ & Katherine A. Brown§

* Division for Protein Structure, National Institute for Medical Research, The Ridgeway, Mill Hill, London, NW7 1AA, UK

† Department of Biochemistry and Genetics, Catherine Cookson Building, New Medical School, University of Newcastle upon Tyne, NE2 4HH, UK

‡ Department of Chemistry, Michigan State University, East Lansing, Michigan 48824-1322, USA

§ Department of Biochemistry, Imperial College of Science, Technology and Medicine, Exhibition Road, London SW7 2AY, UK

Dehydroquinase synthase (DHQS) has long been regarded as a catalytic marvel because of its ability to perform several consecutive chemical reactions in one active site^{1–7}. There has been considerable debate as to whether DHQS is actively involved in all these steps^{1,2}, or whether several steps occur spontaneously, making DHQS a spectator in its own mechanism^{3–5}. DHQS performs the second step in the shikimate pathway, which is required for the synthesis of aromatic compounds in bacteria, microbial eukaryotes and plants⁸. This enzyme is a potential target for new antifungal and antibacterial drugs^{9,10} as the shikimate pathway is absent from mammals and DHQS is required for pathogen virulence¹¹. Here we report the crystal structure of DHQS, which has several unexpected features, including a previously unobserved mode for NAD⁺-binding and an active-site organization that is surprisingly similar to that of alcohol dehydrogenase, in a new protein fold. The structure reveals interactions between the active site and a substrate-analogue inhibitor, which indicate how DHQS can perform multistep catalysis without the formation of unwanted by-products.

DHQS is an NAD⁺-dependent metalloenzyme that converts 3-deoxy-D-arabino-heptulosonate-7-phosphate (DAHP) to dehydroquinase^{1,6}. In microbial eukaryotes, the five central enzymes of the shikimate pathway are encoded on a single pentafunctional polypeptide called AROM, with DHQS forming the amino-terminal domain^{12,13}. The AROM protein is related to two eukaryotic transcription-regulating proteins¹³. We have determined the crystal structure of the DHQS domain of the AROM protein from *Aspergillus nidulans* at 1.8 Å resolution with Zn²⁺, NAD⁺ and a high-affinity inhibitor, carbaphosphonate⁴ (a substrate analogue

of DAHP). The crystal structure shows that the enzyme assembles as a functional homodimer. Each monomer is composed of an N-terminal α/β domain and a carboxy-terminal α -helical domain.

The N-terminal domain includes a Rossmann fold, but surprisingly binds NAD⁺ in an inverted orientation to that observed in all other classic Rossmann fold proteins. This domain consists of a seven-stranded β -sheet, with strand order 1296534 (Fig. 1). Strands β 1 and β 2 form a β -hairpin which is connected by helix α 1 to a five-stranded Rossmann-fold structure^{14,15}, comprising a $\beta\alpha\beta$ unit and a $\beta\alpha\beta\alpha\beta$ unit. Strands β 7 and β 8 form a separate β -hairpin. Both DHQS and the classic Rossmann fold domains bind NAD(P)⁺ at the C-terminal end of the sheet. However, uniquely among known NAD(P)⁺-binding protein structures, the phosphate moieties of NAD⁺ in DHQS are associated with the glycine-rich first turn of the second $\beta\alpha\beta\alpha\beta$ unit, rather than the first ($\beta\alpha$) $\beta\alpha\beta$ unit. This places the nicotinamide moiety on the left of the labelled sheet (β 1 to β 9) in Fig. 1 on the side where the adenosine moiety binds in other NAD⁺-binding proteins. This orientates the NAD⁺ so that the active site is on the opposite side of the sheet to that seen in all other known Rossmann type NAD⁺-binding proteins.

The fold of the C-terminal domain of DHQS shows no detectable structural relationship to known structures^{16–18}. Helix α 6 is surrounded by helices α 7 to α 11 and one 3_{10} helix, forming an antiparallel α -helical barrel. This is followed by a long loop with two short helices (α 12 and α 13), a distorted β -hairpin (β 9 to β 12) and a C-terminal helix α 14. The active site is located in a cleft between the two domains (Fig. 1). The C-terminal domain contains most of the residues involved in catalysis and in substrate and Zn²⁺ binding (Fig. 2). The pentacoordinate Zn²⁺ ion interacts

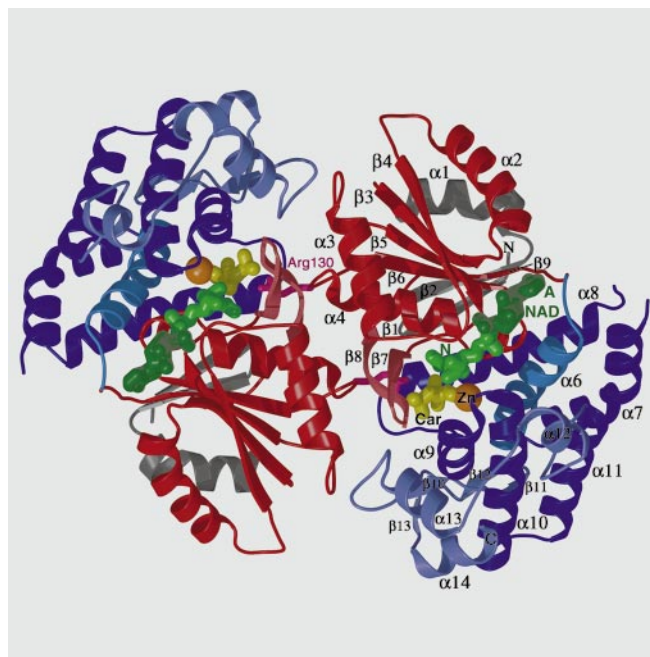


Figure 1 Ribbon diagram of the DHQS dimer showing the inhibitor carbaphosphonate⁴ (yellow), NAD⁺ (green) and Zn²⁺ (orange). The N-terminal β -hairpin and connecting α -helix are shown in grey, the Rossmann fold domain is in red, the β -hairpin insertion is shown in light pink, the C-terminal domain central α -helix is shown in mid-blue, the surrounding five α -helices in dark blue and the C-terminal α -helices and a distorted β -hairpin in grey-blue. The nicotinamide mononucleotide (N) and adenosine (A) moieties of NAD⁺ are shown in light and dark green respectively. The inhibitor carbaphosphonate is labelled Car. Arg 130 is shown in pink (figures produced using BOBSCRIPT²⁰).

with Glu 194, His 271 and His 287 and two inhibitor hydroxyls (Fig. 2).

The complex multistep mechanism of DHQS involves alcohol oxidation, phosphate β -elimination, carbonyl reduction, ring opening and intramolecular aldol condensation¹. It has been proposed that the phosphate catalyses its own elimination^{3,4} and that the final steps in the mechanism could occur spontaneously⁵, and that the enzyme might be required for the prevention of by-product formation in later steps². The structure of DHQS allows the identification of critical interactions that could be used by the enzyme in its proposed catalytic mechanism. The reaction scheme shown in Fig. 3 indicates how interactions between residues and ligands can stabilize intermediates in this complex pathway.

In the first step of the mechanism, the substrate is oxidized at C5 by NAD^+ (Fig. 3). This involves hydride transfer from C5 of DAHP to C4 of the NAD^+ nicotinamide moiety: in the crystal structure the nicotinamide ring is located in a suitable orientation for hydride transfer to occur. In concert, a proton is lost from the C5 hydroxyl group of DAHP. This proton could be relayed to a water molecule (Wat 7), located 2.7 Å from the O5 oxygen of the inhibitor. His 275, adjacent to this water molecule, could then accept this proton in a proton-shuffling system. One of the roles of the Zn^{2+} ion could be to

facilitate hydride transfer and proton loss by polarizing the C5 hydroxyl.

The other NAD^+ -binding enzyme of known structure that has a metal ion bound to the substrate is alcohol dehydrogenase^{19,20}. Like DHQS, this enzyme performs a Zn^{2+} - and NAD^+ -dependent oxidation of alcohol. Comparison of the folds of the catalytic domains of DHQS and alcohol dehydrogenase shows no similarity, and yet the relative positions of active-site components comprising the Zn^{2+} , NAD^+ , substrate alcohol groups and proton-shuffling groups are similar in these enzymes. Thus, this appears to be a case of convergent evolution, where the chemical requirements for alcohol oxidation impose similar local stereochemistry in a completely different protein architecture.

The second step of the mechanism is β -elimination of the phosphate group. The phosphonate oxygens of the inhibitor are bound to Wat 33 and the side chains of Lys 152, Asn 268, His 275 and Lys 356 and Arg 130 from the other monomer (Fig. 3). During phosphate elimination, a proton is removed from C6 of DAHP. Although the phosphate group itself could act as the base in the removal of this proton^{3,4}, the enzyme may catalyse the phosphate elimination by providing a phosphate-binding pocket, formed by Lys 152, Asn 162, Asn 268, His 275 and Lys 356. The structure shows

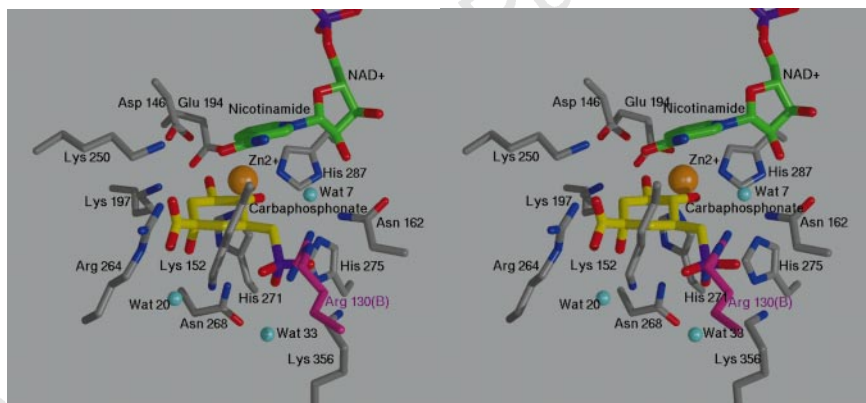


Figure 2 Stereo view of the active site of DHQS. Nitrogen atoms are coloured dark blue, oxygen atoms red and phosphorus atoms purple. Carbon atoms are coloured yellow in the inhibitor carbaphosphonate, green in NAD^+ , grey in chain A

protein side chains and pink in the side chain of Arg 130 from chain B. Water molecules and the Zn^{2+} ion are represented by pale blue and orange spheres, respectively.

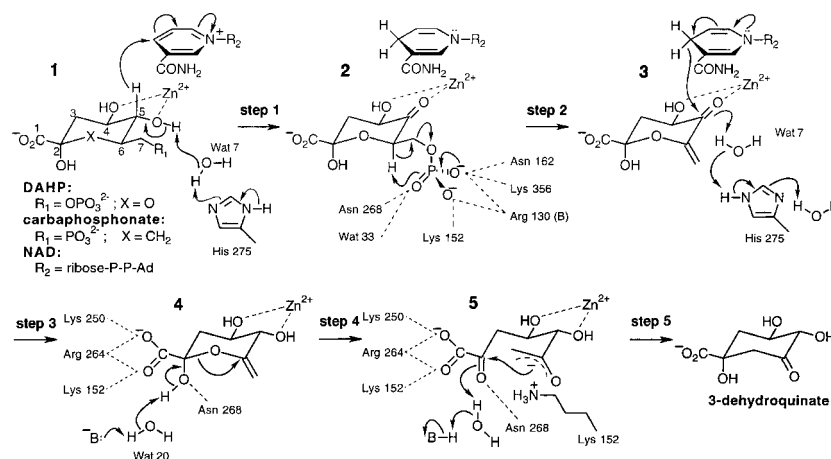


Figure 3 The proposed mechanism for conversion of DAHP to dehydroquinate by DHQS. Step 1 is oxidation of the hydroxyl group attached to C5; step 2 is β -elimination of inorganic phosphate; step 3 is reduction of the ketone at C5; step 4

is a ring opening, and step 5 is an intramolecular aldol condensation. Ad, adenosine; P, phosphate; B, a general base. Lys 152 is shown fulfilling two possible roles.

Table 1 Structure determination and refinement statistics

Dataset	Temp (K)	Cell parameters <i>a</i> (Å)	<i>b</i> (Å)	<i>c</i> (Å)	Resolution limits (Å)	Completeness (last shell) (%)	Total Reflections	Unique	<i>R</i> _{syn} (%)* (last shell)	(<i>I</i> /σ(<i>I</i>)) (last shell)
Form A	277	69.0	80.9	144.7	15.0–3.0	99.7 (100.0)	65,925	16,566	8.1 (17.7)	16.1 (7.6)
Form A	100	67.5	80.7	143.3	20.0–2.3	92.3 (94.5)	123,775	32,698	8.9 (28.4)	13.6 (4.5)
Form A	100	67.7	80.8	143.5	25.0–1.7	91.2 (84.8)	297,189	79,632	6.8 (34.1)	15.1 (3.2)
Hg derivative	277	69.4	81.6	144.9	12.0–3.6	95.9 (96.9)	43,387	9,429	10.2 (19.1)	13.8 (9.6)
Form B	277	65.9	71.3	200.1	20.0–2.8	88.9 (92.0)	95,653	21,195	94. (35.6)	13.6 (3.5)
HgOAc derivative statistics										
Number of sites		<i>R</i> _{iso} (%)†		Phasing power‡		Overall FOM§	<i>R</i> _{culis}	Anomalous <i>R</i> _{culis} ¶		
8		21.6		1.80		0.41	0.60	0.94		
Refinement statistics										
Resolution limits (Å)	Atoms	residues	solvent	<i>R</i> -value# (%)	Free- <i>R</i> value☆ (%)	r.m.s. deviations from ideal values**				Average <i>B</i> value†† (Å ²)
25.0–1.8	6539	761	718	17.3	22.4	Bond lengths (Å)	Bond angles (°)			20.6
						0.013	2.3			

* $R_{\text{syn}} = 100 \times \sum_{hkl} \{ \sum_i |I_i(hkl)| - \langle I_i(hkl) \rangle \} / \sum_i |I_i(hkl)|$, where $I_i(hkl)$ are the intensities of multiple measurements and $\langle I_i(hkl) \rangle$ is the average of the measured intensities of the i th reflection, summing over all the measured reflections.

† $R_{\text{iso}} = 100 \times \sum_{hkl} (|F_{\text{PH}} - F_{\text{P}}|) / \sum_{hkl} |F_{\text{P}}|$, where F_{PH} is the derivative structure factor and F_{P} is the native structure factor and the sum is performed over all reflections that occur in the two data sets.

|| $R_{\text{culis}} = 100 \times \sum \{ |F_{\text{PH}}(hkl)| \pm |F_{\text{P}}(hkl)| - |F_{\text{H}}(hkl)| / \sum \{ |F_{\text{PH}}(hkl)| \pm |F_{\text{P}}(hkl)| \}$, where the sum is over all the centric reflections.

‡ Phasing power = $\langle F_{\text{H}} \rangle / \epsilon$, where ϵ is the lack of closure error.

§ Overall FOM is the overall figure of merit.

¶ Anomalous $R_{\text{culis}} = 100 \times \sum \{ |F_{\text{PH}}(hkl)| \pm |F_{\text{P}}(hkl)| - |F_{\text{H}}(hkl)| / \sum \{ |F_{\text{PH}}(hkl)| \pm |F_{\text{P}}(hkl)| \}$, where the sum is over all the centric reflections.

Crystallographic *R* value = $(\sum_{hkl} |F_{\text{O}} - F_{\text{C}}|) / \sum_{hkl} |F_{\text{O}}|$, where F_{O} and F_{C} are the observed and calculated structure factors, respectively, summed over the 95% of the native data used for the refinement.

☆ Free *R* value = $(\sum_{hkl} |F_{\text{O}}| - |F_{\text{C}}|) / \sum_{hkl} |F_{\text{O}}|$, summed over the 5% of the native data excluded from the refinement.

** R.m.s. is the square root of the mean of the squares of the deviations of the bond lengths and angles from their ideal values.

†† Average *B* value summed over all atoms.

that these interactions could force the phosphate oxygens into a position where they can remove the proton from C6. The enzyme may also selectively stabilize an E1cb-like transition state or reactive intermediate⁷.

In the third step, the C5 ketone of intermediate **3** in Fig. 3 is reduced by NADH. This is a reversal of the first step, with hydride ion from C4 of NADH being transferred back to C5 of the substrate. Using the proton-shuffling system, a proton could be transferred to the ketone oxygen from His 275, via Wat 7, creating intermediate **4** (Fig. 3).

The final two steps of the reaction consist of ring opening and intramolecular aldol condensation (Fig. 3). The ring-opening step involves deprotonation of the hydroxyl at C2. The crystal structure indicates that Wat 20 could participate in this step because the only nearby side chain, Asn 268, is unlikely to act as a proton acceptor (Fig. 2). The side chain of Arg 260 could act as the base that accepts a proton from Wat 20. Ring opening is followed by a rotation about the C5–C6 bond and an intramolecular aldol condensation to reform the ring. These two steps can occur spontaneously in solution when intermediate **4** is generated from a photo-labile precursor⁵. In solution a minor by-product, a C2 epimer of 3-dehydroquinone, is also produced which is not observed in the enzyme-catalysed reaction². This crystal structure shows that epimerization at C2 would be prevented by the interactions between the carboxylic acid group on C2 of intermediate **4** and the side chains of Lys 152, Lys 250 and Arg 264, and by the interactions between the hydroxyl or ketone on C2 with Asn 268 and Wat 20. The hydroxyl groups on C4 and C5 interact with the Zn²⁺ ion and the C4 hydroxyl also interacts with the side chains of Asp 146, Glu 194 and Lys 197. Rotation about the C5–C6 bond is possible, however. The negative charge on intermediate **5** (Fig. 3) could be stabilized by interactions with a nearby side chain such as Lys 152. Alternatively, the enolate anion in intermediate **5** could be temporarily protonated to form an enol, perhaps by donation of a proton from the side chain of Lys 152. The protein therefore binds intermediate **5** to prevent the formation of side products, as proposed earlier².

The crystal structure of DHQS reveals an activate site possessing a rich variety of interactions between metal, dinucleotide, substrate analogue and protein. The mechanism therefore involves a complex orchestrated interplay between all of these structural elements. The active-site components involved in the redox steps of the reaction provide a classic example of convergent evolution with alcohol

dehydrogenase. Other interactions between the substrate analogue and the enzyme indicate that the enzyme is much more than a simple oxidoreductase. These structural and mechanistic observations provide a detailed framework for the design of new antimicrobial compounds. Finally, the arrangement of the active site of DHQS indicates that the enzyme is not just a spectator in catalysis but stabilizes intermediates and prevents side reactions through its entire reaction pathway. □

Methods

Data collection and reduction. Expression²¹, purification²² and crystallization of DHQS are to be described elsewhere (K.A.B. et al., unpublished results). Two orthorhombic crystal forms, A and B, in space group *P*₂₁₂₁₂, were obtained for the complex between DHQS, NAD⁺ and carbaphosphonate⁴ (Table 1). Data were collected either with a rotating-anode source and an RAXISII detector, or at the SRC, CCLRC Daresbury laboratory, UK, on beamlines 7.2, 9.5 or 9.6. A single mercuric acetate derivative was obtained by soaking a form-A crystal for 3 h in 0.5 mM mercurous acetate in mother liquor and data were collected using a wavelength of 0.87 Å. Data were integrated, scaled and reduced using DENZO²³ and SCALEPACK²³; the CCP4 suite of crystallographic programs²⁴ was used for subsequent steps.

Single isomorphous replacement phasing. Heavy-atom sites were identified with the program HASSP²⁵, Patterson maps and difference Pattersons. Initial phases from eight Hg sites and anomalous data were refined using MLPHARE²⁴.

Density modification and refinement. Six-fold averaging (with the dimers in the two form-A and one form-B datasets) and solvent flattening using DMULTI²⁴ gave an interpretable map. The structure was built using the program O (ref. 26) in a map generated from the 3-Å dataset; the model was then extensively rebuilt using the 25.0–1.8 Å data from the 1.7-Å dataset. Refinement proceeded using TNT²⁷ and REFMAC²⁴, with a total of 64,534 reflections (3,385 reflections were excluded from the refinement for cross-validation²⁸). Towards the end of refinement, the non-crystallographic symmetry restraints were removed. Electron density was visible for residues A4 to A223, A231 to A391, B3 to B223, B235 to B393.

Received 25 February; accepted 6 May 1998.

- Srinivasan, P. R., Rothchild, J. & Sprinson, D. B. The enzymatic conversion of 3-deoxy-D-arabino-heptulosinic acid 7-phosphate to 5-dehydroquinone. *J. Biol. Chem.* **238**, 3176–3182 (1963).
- Bartlett, P. A., McLaren, K. L. & Marx, M. A. Divergence between the enzyme-catalysed and noncatalysed synthesis of 3-dehydroquinone. *J. Org. Chem.* **59**, 2082–2085 (1994).
- Widlanski, T. S., Bender, S. L. & Knowles, J. R. Dehydroquinone synthase: A sheep in wolf's clothing? *J. Am. Chem. Soc.* **111**, 2299–2300 (1989).
- Bender, S. L., Widlanski, T. & Knowles, J. R. Dehydroquinone synthase: The use of substrate analogues to probe the early steps of the catalysed reaction. *Biochemistry* **28**, 7560–7572 (1989).

5. Bartlett, P. A. & Satake, K. Does dehydroquinase synthesize dehydroquinone? *J. Am. Chem. Soc.* **110**, 1628–1630 (1988).
6. Lambert, J. M., Boocock, M. R. & Coggins, J. R. The 3-dehydroquinase synthase activity of the pentafunctional arom enzyme complex of *Neurospora crassa* is Zn^{2+} dependent. *Biochem. J.* **226**, 817–829 (1985).
7. Montchamp, J.-L. & Frost, J. W. Cyclohexenyl and cyclohexylidene inhibitors of 3-dehydroquinase synthase: Active site interactions relevant to enzyme mechanisms and inhibitor design. *J. Am. Chem. Soc.* **119**, 7645–7653 (1997).
8. Bentley, R. The shikimate pathway—a metabolic tree with many branches. *Crit. Rev. Biochem. Mol. Biol.* **25**, 307–384 (1990).
9. Kishore, G. M. & Shah, D. M. Amino acid biosynthesis inhibitors as herbicides. *Annu. Rev. Biochem.* **57**, 27–63 (1988).
10. Pittard, A. J. Biosynthesis of aromatic amino acids. *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology* (eds Niedhardt, F. C. et al.) **1**, 368–394 (Am. Soc. Microbiol., Washington, DC, 1987).
11. Günel-Özcan, A., Brown, K. A., Allen, A. & Maskell, D. J. *Salmonella typhimurium aroB* mutants are attenuated in BALB/c mice. *Microbial Pathogen.* **17**, 169–174 (1997).
12. Case, M. E. & Giles, N. H. Partial enzyme aggregates formed by pleiotropic mutants in the arom gene cluster of *Neurospora crassa*. *Proc. Natl. Acad. Sci. USA* **68**, 58–62 (1971).
13. Hawkins, A. R., Lamb, H. K., Moore, J. D., Charles, I. G. & Roberts, C. F. The prechorismate (shikimate) and quinate pathways in filamentous fungi: theoretical and practical aspects. *J. Gen. Microbiol.* **139**, 2891–2899 (1993).
14. Rossmann, M. G., Moras, D. & Olsen, K. W. Chemical and biological evolution of a nucleotide-binding protein. *Nature* **250**, 194–199 (1974).
15. Lesk, A. M. NAD-binding domains of dehydrogenases. *Curr. Opin. Struct. Biol.* **5**, 775–783 (1995).
16. Kleywegt, G. J. & Jones, T. A. Where freedom is given, liberties are taken. *Structure* **3**, 535–540 (1995).
17. Holm, L. & Sander, C. Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123–138 (1993).
18. Murzin, A. G., Brenner, S. E., Hubbard, T. & Chothia, C. SCOP-A structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* **247**, 536–540 (1995).
19. Eklund, H. et al. Three-dimensional structure of horse liver alcohol dehydrogenase at 2.4 Å resolution. *J. Mol. Biol.* **102**, 27–59 (1976).
20. Eklund, H. et al. Structure of the triclinic ternary complex of horse liver alcohol dehydrogenase at 2.9 Å resolution. *J. Mol. Biol.* **146**, 561–587 (1981).
21. Van Den Hombergh, J. P. T. W., Moore, J. D., Charles, I. G. & Hawkins, A. R. Overproduction in *Escherichia coli* of the dehydroquinase synthase domain of the *Aspergillus nidulans* pentafunctional AROM protein. *Biochem. J.* **284**, 861–867 (1992).
22. Moore, J. D., Coggins, J. R., Virden, R. & Hawkins, A. R. Efficient independent activity of a monomeric, monofunctional dehydroquinase synthase derived from the N-terminus of the pentafunctional AROM protein of *Aspergillus nidulans*. *Biochem. J.* **301**, 297–304 (1994).
23. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Meth. Enzymol.* **276**, 307–326 (1996).
24. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
25. Terwilliger, T. C., Kim, S.-H. & Eisenberg, D. Generalised method of determining heavy atom positions using the difference Patterson function. *Acta Crystallogr. A* **43**, 1–5 (1987).
26. Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron-density and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
27. Tronrud, D. E., *The TNT Refinement Package Manual*, 1–191, Oregon State Board of Higher Education (1993).
28. Brunger, A. T. Free *R*-value—a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472–475 (1992).
29. Esnouf, R. M. An extensively modified version of Molscript that includes greatly enhanced coloring capabilities. *J. Mol. Graph. Model.* **15**, 132 (1997).

Acknowledgements. We thank the support staff at the Synchrotron Radiation Source at Daresbury Laboratory, UK, for assistance; D. R. Swatman, E.T.F. Tsai, R. R. Patel and Y. S. Li for technical assistance; G. G. Dodson for support and encouragement; K. Henrick, S. Gamblin, M. Hirschberg, J. O. Baum, W. Taylor and J. D. Moore for discussion; and P. Bartlett for helpful comments. This work was supported by BBSRC. E.P.C. was supported by the UK MRC and K.A.B. received a BBSRC Advanced Fellowship.

Correspondence and requests for materials should be addressed to K.A.B. (e-mail: k.brown@ic.ac.uk). Coordinates have been deposited with the Brookhaven Databank under accession number 1dqs.

addendum

A prolactin-releasing peptide in the brain

Shuji Hinuma, Yugo Habata, Ryo Fujii, Yuji Kawamata, Masaki Hosoya, Shoji Fukusumi, Chieko Kitada, Yoshinori Masuo, Tsuneo Asano, Hirokazu Matsumoto, Masahiro Sekiguchi, Tsutomu Kurokawa, Osamu Nishimura, Haruo Onda & Masahiko Fujino

Nature **393**, 272–276 (1998)

The prolactin-releasing peptide cDNA sequence data have been submitted to the DDBJ/EMBL/GenBank databases. The accession numbers are as follows: AB015417, *Bos taurus* mRNA for preprolactin-releasing peptide; AB015418, *Rattus norvegicus* mRNA for preprolactin-releasing peptide; AB015419, *Homo sapiens* mRNA for preprolactin-releasing peptide. □

correction

Engineering cyclophilin into a proline-specific endopeptidase

Eric Quéméneur, Mireille Moutiez, Jean-Baptiste Charbonnier & André Ménez

Nature **391**, 301–304 (1998)

The efficiency value (k_{cat}/K_m) of cyproase 1 is equal to $0.2 \times 10^4 \text{ M}^{-1} \text{ s}^{-1}$, and not to $0.7 \times 10^4 \text{ M}^{-1} \text{ s}^{-1}$ as published. Also, the histidine at residue 104 titrates with a pK equal to 6.47 ± 0.16 and 6.74 ± 0.15 when measuring K_{cat} and the k_{cat}/K_m , respectively, and not to 6.74 ± 0.16 and 6.74 ± 0.15 . □